JULY/AUGUST 2013 • VOLUME 84, NUMBER 4

CHILD DEVELOPMENT

Editor
JEFFREY J. LOCKMAN

Associate Editors
KAREN L. BIERMAN • MARGARET BURCHINAL • JOHN COLOMBO
MICHAEL CUNNINGHAM • GIL DIESENDRUCK • ANDREW J. FULIGNI
MARY GAUVAIN • ERIKA HOFF • MELANIE KILLEN • BONNIE J. LEADBEATER
LOUIS J. MOSES • GLENN I. ROISMAN • CYNTHIA STIFTER

PRIOR 10
Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

SRCD
Society for Research
in Child Development

Published by Wiley for the Society for Research in Child Development

Discover this journal online at
WILEY ONLINE LIBRARY
wileyonlinelibrary.com/journal/cdev

Published by Wiley for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2013 are: Print & Online: US$836 (US), US$1170 (Rest of World), €761 (Europe), £598 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2009, where available. For other pricing options, including access information and terms and conditions, please visit www.wileyonlinelibrary.com/access.

Delivery Terms and Legal Title: Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability.

For more information about Wiley journals, including online access information, terms and conditions, and other pricing options, please visit wileyonlinelibrary.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2013 includes six issues of *Monographs of the Society for Research in Child Development* and four issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at wileyonlinelibrary.com/journal/cdev. This journal is available online at Wiley Online Library. Visit wileyonlinelibrary.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the AGORA initiative with the FAO, the HINARI initiative with the WHO and the OARE initiative with UNEP. For information, visit www.aginternetwork.org, www.healthinternetwork.org, www.oaresciences.org.

Journal Customer Services: For ordering information, claims and any inquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office.

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada).

Europe, Middle East and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315.

Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000.

Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.

Visit our Online Customer Get-Help available in 6 languages at www.wileycustomerhelp.com

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Child Development © 2013 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley for the Society for Research in Child Development

Child Development

Editor

Jeffrey J. Lockman
Tulane University

Associate Editors

Karen L. Bierman, *The Pennsylvania State University*
Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Gil Diesendruck, *Bar-Ilan University*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Erika Hoff, *Florida Atlantic University*
Melanie Killen, *University of Maryland*
Bonnie J. Leadbeater, *University of Victoria*
Louis J. Moses, *University of Oregon*
Glenn I. Roisman, *University of Minnesota*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor

Adam Martin

Editorial Assistant

Stephanie DeFouw

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
Christa Anderson
William F. Arsenio
Daniel H. Ashmead
Karen Bartsch
Jay Belsky
Aprile D. Benner
Annie Bernier
Ramesh Bhatt
Susan Birch
Clancy Blair
Amy Booth
Gregory Stephen Braswell
Mara Brendgen
Keith Burt
Emily W. Bushnell
Susan B. Campbell
Noel Card
Stephanie M. Carlson
Cara Cashon
Pablo Chavajay
Rebekah Coley
Catherine Costigan
Elysia Davis
Jed Elison
Ruth Feldman
Lisa Gatzke-Kopp
Heidi Gazelle
Abi Gewirtz
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Dan Hart
Wendy Hoglund
Amy Joh
Roger Kobak
Ann Kruger
Kristin H. Lagattuta
Charlie Lewis
Ramaswami Mahalingam
Tina Malti
Jelani Mandara
Virginia Marchman
Ann S. Masten
Clark McKown
Bob McMurray
Jayanthi Mistry
Ginger A. Moore
Marelene Moretti
Dianna Murray-Close
Velma McBride Murry
Letitia Naigles
Jenae Neiderhiser
Drew Nesdale
Nora S. Newcombe
Mark Nielsen
Candice L. Odgers
Anna Papafragou
Kristin Pauker
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Samuel Putnam
Robert Reeve
Greg Reynolds
Marjorie Rhodes
John J. Rieser
Kristen Ritchey
Deborah Rivas-Drake
Phil Rodkin
Susan A. Rose
Martin D. Ruck
Adam Rutland
David Schwartz
Eleanor Seaton
Stefanie Sinno
Judi G. Smetana
Catherine Snow
Dena Phillips Swanson
Moin Syed
Adriana Umaña-Taylor
Rebecca Treiman
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Peter Vishton
Elizabeth Votruba-Drzal
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, July/August 2013, Volume 84, Number 4, Pages 1115–1490

Contents

IN THIS ISSUE 1115

REVIEW

Social Equity Theory and Racial-Ethnic Achievement Gaps
Clark McKown 1120

EMPIRICAL REPORTS

Local Redundancy Governs Infants' Spontaneous Orienting to Visual-Temporal Sequences
Caspar Addyman and Denis Mareschal 1137

Longitudinal Relations of Intrusive Parenting and Effortful Control to Ego-Resiliency During Early Childhood
Zoe E. Taylor, Nancy Eisenberg, Tracy L. Spinrad, and Keith F. Widaman 1145

EMPIRICAL ARTICLES

Little Evidence That Time in Child Care Causes Externalizing Problems During Early Childhood in Norway
Henrik D. Zachrisson, Eric Dearing, Ratib Lekhal, and Claudio O. Toppelberg 1152

Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics
Tran D. Keys, George Farkas, Margaret R. Burchinal, Greg J. Duncan, Deborah L. Vandell, Weilin Li, Erik A. Ruzek, and Carollee Howes 1171

Child Care in Infancy and Cognitive Performance Until Middle Childhood in the Millennium Cohort Study
Sylvana M. Côté, Orla Doyle, Amélie Petitclerc, and Lori Timmins 1191

Genetic Moderation of Early Child-Care Effects on Social Functioning Across Childhood: A Developmental Analysis
Jay Belsky and Michael Pluess 1209

Coparenting Among Lesbian, Gay, and Heterosexual Couples: Associations With Adopted Children's Outcomes
Rachel H. Farr and Charlotte J. Patterson 1226

Role Models and the Psychological Characteristics That Buffer Low-Socioeconomic-Status Youth From Cardiovascular Risk
Edith Chen, William K. Lee, Lisa Cavey, and Amanda Ho 1241

Informants' Traits Weigh Heavily in Young Children's Trust in Testimony and in Their Epistemic Inferences
Jonathan D. Lane, Henry M. Wellman, and Susan A. Gelman 1253

Rigidity in Gender-Typed Behaviors in Early Childhood: A Longitudinal Study of Ethnic Minority Children
May Ling Halim, Diane Ruble, Catherine Tamis-Lemonda, and Patrick E. Shrout 1269

Forces and Motion: How Young Children Understand Causal Events
Tilbe Göksun, Nathan R. George, Kathy Hirsh-Pasek, and Roberta M. Golinkoff 1285

Sociocultural Settings Influence the Emergence of Prelinguistic Deictic Gestures
Dorothe Salomo and Ulf Liszkowski 1296

Dual Routes to Cognitive Flexibility: Learning and Response-Conflict Resolution in the Dimensional Change Card Sort Task
Michael Ramscar, Melody Dye, Jessica W. Gustafson, and Joseph Klein 1308

Increases in Cognitive and Linguistic Processing Primarily Account for Increases in Speaking Rate With Age
Ignatius S. B. Nip and Jordan R. Green 1324

Associations Among Pubertal Development, Empathic Ability, and Neural Responses While Witnessing Peer Rejection in Adolescence
Carrie L. Masten, Naomi I. Eisenberger, Jennifer H. Pfeifer, Natalie L. Colich, and Mirella Dapretto 1338

Developmental Trajectories of Acculturation in Hispanic Adolescents: Associations With Family Functioning and Adolescent Risk Behavior
Seth J. Schwartz, Sabrina Des Rosiers, Shi Huang, Byron L. Zamboanga, Jennifer B. Unger, George P. Knight, Hilda Pantin, and José Szapocznik 1355

Associations Between First-Time Expectant Women's Representations of Attachment and Their Physiological Reactivity to Infant Cry
Jennifer C. Ablow, Amy K. Marks, S. Shirley Feldman, and Lynne C. Huffman 1373

Visual Attention to Global and Local Stimulus Properties in 6-Month-Old Infants: Individual Differences and Event-Related Potentials
Maggie W. Guy, Greg D. Reynolds, and Dantong Zhang 1392

The Moving Window Technique: A Window Into Developmental Changes in Attention During Facial Emotion Recognition
Elina Birmingham, Tamara Meixner, Grace Iarocci, Christopher Kanan, Daniel Smilek, and James W. Tanaka 1407

Daily Intragroup Contact in Diverse Settings: Implications for Asian Adolescents' Ethnic Identity
Tiffany Yip, Sara Douglass, and J. Nicole Shelton 1425

Siblings, Theory of Mind, and Executive Functioning in Children Aged 3–6 Years: New Longitudinal Evidence
Anna R. McAlister and Candida C. Peterson 1442

"Two for Flinching": Children's and Adolescents' Narrative Accounts of Harming Their Friends and Siblings
Holly Recchia, Cecilia Wainryb, and Monisha Pasupathi 1459

Predicting Long-Term Growth in Students' Mathematics Achievement: The Unique Contributions of Motivation and Cognitive Strategies
Kou Murayama, Reinhard Pekrun, Stephanie Lichtenfeld, and Rudolf vom Hofe 1475

OTHER

Manuscripts Accepted

For the complete list of Manuscripts Accepted for Publication, please visit www.srcd.org and navigate to ***Child Development*** *in the Publications section.*

Child Development, July/August 2013, Volume 84, Number 4, Pages 1115–1119

In This Issue

Race and Ethnicity

In the United States, children from different racial and ethnic groups tend to get different results on tests of school readiness and academic outcomes, creating racial-ethnic achievement gaps that begin early and last through all levels of school. **McKown** (p. 1120) proposes Social Equity Theory, which says that racial-ethnic achievement gaps come from two kinds of social processes—direct influences (social processes that support achievement) and signal influences (cues that communicate negative expectations about a child's racial-ethnic group)—that operate across key developmental contexts. These two influences combine to create racial-ethnic achievement gaps that vary by setting and age. The theory can inform interventions to reduce the achievement gap.

Adolescents living in diverse, urban areas come into contact with people of their own and different races daily. **Yip, Douglass, and Shelton** (p. 1425) asked how these interactions influence how about 130 Asian youths (mean age of 14, primarily from a mix of working- and middle-class families) feel about their own ethnic group and whether the racial composition of their school makes a difference in terms of their feelings about their own ethnic group (some students went to primarily Asian schools, some attended predominantly White schools, and some went to racially mixed schools). Findings: When Asian adolescents were in the numerical minority in their school, contact with other Asian people was associated with feeling better about one's ethnic heritage, particularly for those teens with a strong ethnic identity; this feeling carried over to the next day. The findings can help us develop a richer and more nuanced understanding of the role of ethnicity and race in adolescents' daily lives.

Perspectives on Parenting

Intrusive parenting is negatively related to children's personality resiliency through its negative effects on children's ability to regulate attention and behavior. That's what **Taylor, Eisenberg, Spinrad, and Widaman** (p. 1145) find in their longitudinal study of relationships between parenting behaviors and early temperamental traits. They looked at more than 250 mostly White, middle-class families when children were 18, 30, and 42 months old. At 18 months, intrusive parenting was negatively associated with children's adaptability to environmental stress and change, while the ability to focus and shift attention was positively associated with this type of adaptability, both concurrently and across time. The findings suggest that interventions that focus on reducing negative parental behaviors may affect not only children's regulatory abilities but also their skills at adapting to stressful situations.

Regardless of parents' sexual orientation, adopted children's outcomes are linked to how well parents work together. That's what **Farr and Patterson** (p. 1226) find in their examination of associations between coparenting and children's behavior among about 100 mostly White and well-educated adoptive families headed by lesbian, gay, or heterosexual couples with preschoolers. Lesbian and gay couples were likely to share child-care tasks, while heterosexual couples were likely to specialize, with more moms caring for the children than dads. This suggests that lesbian and gay couples are creating new ways to live together and raise children outside of traditional gender roles. The best predictors of children's behavior problems were competition between parents and dissatisfaction with how child care was divided—neither of which differed by parents' sexual orientation.

Considering Child Care

More hours in child care doesn't lead to more aggression and noncompliance in children, according to **Zachrisson, Dearing, Lekhal, and Toppelberg** (p. 1152), based on their study out of Norway, which has universal access to subsidized center care with regulated quality standards. This finding contrasts with U.S. findings, which have indicated that children who spend more time in child care have, on average, higher levels of aggression and noncompliance than children in little or no child care. One reason for this difference may be that Norway has a comprehensive early childhood policy that

© 2013 The Author
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0001
DOI: 10.1111/cdev.12146

also includes a year of paid parental leave for the child's first year, so children in Norway start child care as toddlers, while many children in the United States enter child care as infants. The study followed 72,000 mothers and their children, including 8,000 sibling pairs, with a focus on mothers' reports of externalizing problems at 18 and 36 months, and on the hours children were in child care at the same time points; most children were in child care for 40 hours a week or less.

Many children spend numerous hours during preschool in child-care settings, and the quality of their care is believed to be an important factor in their development. **Keys et al.** (p. 1171) used data from four large studies (totaling 6,250 children) and meta-analysis techniques to predict the effects of quality of preschool center care on children's school readiness skills. Across a one- to two-school-year period, children who experienced higher quality preschool care showed modest gains in language and mathematics skills, but not in social skills and problem behaviors. This suggests that efforts to improve school readiness skills by boosting the quality of child care in the preschool years will have only modest success.

Most children in developed countries receive some form of child care during early childhood. **Côté, Doyle, Petitclerc, and Timmins** (p. 1191) examined the effect of child care on the cognitive outcomes of 13,000 children at ages 3, 5, and 7, using the British Millennium Cohort Study. Receiving any type of child care up to 9 months was associated with better cognitive outcomes at age 3, but only for children of mothers with low education; these effects did not persist to ages 5 or 7. Early center-based (or formal) child care was associated with better cognitive outcomes than informal care at ages 3 and 5 years, but not at 7. Furthermore, the impact of child care did not differ based on family income. The results suggest that social programs that provide child care to vulnerable populations should target low maternal education rather than low family income.

The effects of the quality of child care depend on a child's genetic makeup, according to a longitudinal study by **Belsky and Pluess** (p. 1209) that's the first to address this question. The study focused on two genetic markers in more than 500 White children who took part in the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development and three different features of child care (quality, quantity, and type) to predict children's behavior problems and social competence from age 4 through sixth grade. Findings: Children who carried one version of a genotype had more behavior problems and less social competence when in low-quality care, and more social competence when in high-quality care, although the effects faded over time. The study suggests that failing to consider children's genetic makeup may lead to inaccurate estimates of the effects of child care because not all children are equally susceptible to, in this case, the benefits of high-quality care and the costs of low-quality care with respect to their social development.

Benefits of Role Models

Children who are raised in low-income homes often have poor health outcomes. In their study of about 160 youths (ages 13–16) from a variety of socioeconomic backgrounds, **Chen, Lee, Cavey, and Ho** (p. 1241) asked about role models, coping methods, and beliefs about the future, and took blood to assess inflammatory markers that predict cardiovascular disease. They find a biological benefit—less inflammation—for low-socioeconomic status youths who had supportive role models and also for those who engaged in adaptive coping like "shift-and-persist" strategies that reframe stressful events more positively. These strategies may begin to reduce the physiological burden of growing up in low-income neighborhoods.

Biases in Thinking: Traits and Gender

People's traits influence how young children learn from them, according to **Lane, Wellman, and Gelman** (p. 1253). They examined how people's traits—specifically their smartness, kindness, and honesty—affect how 3- to 6-year-olds seek information from people, trust their claims, and make judgments about those people's knowledge. They studied about 80 primarily White, middle-class children and two dozen adults to find that children (and adults) clearly tracked people's traits when learning from them, preferring to ask and endorse information provided by people who were nice, smart, and honest. And these traits influenced the knowledge that young children attributed to the informants. They find that even though young preschoolers understand what it means to be knowledgeable and ignorant, they inaccurately over-attribute knowledge to people who are nice, smart, and honest.

While preschoolers tend to act in highly gender-stereotypical ways (think girls dressing up as princesses), by age 5, they may begin to reject gender stereotypes and be more flexible in following gender norms. Most research in this area has been done with middle-income, ethnic-majority groups. In their longitudinal study, **Halim, Ruble, Tamis-LeMonda, and Shrout** (p. 1269) pursued the developmental course of gender-stereotypic behaviors in low-income, ethnic-minority children, asking whether change in children's gender-stereotypic behaviors is due to children's growing awareness of being a boy or girl and their accompanying active search for and acquisition of knowledge of gender stereotypes during this period. Almost 230 Mexican, Dominican, and African American children were examined annually. From ages 3 to 4, children increasingly played dress-up, dressed in feminine and masculine ways, and tended to have more same-gender friends. For certain behaviors, children acted in less gender-stereotypical ways at age 5, indicating movement toward flexibility, with timing and levels of behavior varying by ethnic group.

Forces of Nature

Can children predict cause–effect relations within events such as throwing Frisbees and blocking soccer balls? **Göksun, George, Hirsh-Pasek, and Golinkoff** (p. 1285) looked at whether preschool children understand the relation between forces and motion in causal events, specifically whether young children can integrate how two forces affect an object's movement. Their study of 60 primarily White middle-class children ages 3–5 finds that the children understood the forces in causal events only incompletely, with understanding of more complex relations developing around kindergarten age. By providing evidence about how preschoolers use force and motion to evaluate the causal relations they witness, the study suggests that children may not fully understand complex causal chains until later than previously thought.

Contextual Supports

Infants learn words best when their caregivers use the words in an ongoing activity with the infants and when caregivers gesture to the objects they're talking about. However, before they can speak, infants already communicate with gestures to direct attention to things they like or want. And the more infants gesture, the earlier they begin to speak. **Salomo and Liszkowski** (p. 1296) looked at the daily activities of about 50 infants (ages 8–15 months) in three settings—Yucatec Mayans in Mexico, Dutch in the Netherlands, and Mandarin-Chinese in Shanghai. They find that the infants were immersed in social activity related to various objects to varying degrees and received different amounts of gestures based on the culture they came from. Critically, the amount of gestures infants received correlated with their own production of such gestures. This supports the idea that infants' preverbal gestures, like later first words, emerge through social activities and the gestures directed to them.

Why can many preschoolers transition from playtime to quiet time without incident, but fail to shift from playing by one rule to using another in simple cognitive tests? **Ramscar, Dye, Gustafson, and Klein** (p. 1308) sought to learn why very young children seem capable of engaging in nuanced and context-appropriate behavior in certain situations but not others. In their study of almost fifty 3-year-olds from a range of socioeconomic and racial/ethnic backgrounds, they find that when children are given the right contextual information, they can succeed in a test of cognitive control at which they would otherwise fail. This supports the idea that there are dual routes to smart, flexible behavior, and has theoretical implications both for how we conceptualize the developing mind and for how we can exploit the principles of learning theory to help children navigate tasks they might otherwise find difficult.

Speak Easy

Children's rate of speech gradually and steadily increases during the first two decades of life. Research on the factors that allow children to talk faster as they grow older is important for understanding normal speech development and how delays in physical growth, motor function, language, and cognition may impede it. **Nip and Green** (p. 1324) studied about 50 primarily White, middle-class children and adults ranging in age from 4 to 24 to learn about changes in speaking rates, using an innovative optical motion capture system to record mouth movements during speech. They find that increases in speaking rate as children age are driven primarily by faster language and cognitive processing rather than by factors related to physical maturation or growth. The findings can inform efforts to identify and treat children with speech delays and difficulties.

Teens: Peer Rejection and Acculturation

Even when teens avoid being the actual target of peer rejection and bullying, they're negatively affected by witnessing these behaviors. **Masten, Eisenberger, Pfeifer, Colich, and Dapretto** (p. 1338) studied sixteen 13-year-olds who watched peers playing a game in which one player was excluded; the participants, from diverse ethnic and socioeconomic backgrounds, also completed questionnaires to gauge how various changes in physical and emotional development related to brain activity while they watched. Teens who showed maturity in different aspects of physical and emotional development—those who had undergone more rapid development in puberty—had stronger activity in brain regions known to be associated with taking others' perspectives. Teens with more ability to empathize with others' distress showed stronger activity in brain regions involved in processing distress and pain. In suggesting that teens at different stages and rates of development respond differently to peer rejection, this study can inform parents and teachers in their efforts to reduce the negative impact of bullying.

Hispanic immigrants have been coming to the United States in large numbers since the 1980s, and studies suggest that acculturation can predispose adolescents toward aggressive or delinquent behavior, drug and alcohol use, and early and unsafe sexual activity. **Schwartz et al.** (p. 1355) took a longitudinal look at patterns of cultural adjustment over 2 years (between seventh and ninth grades) in about 270 Hispanic immigrants; the researchers also interviewed the youths' parents, most of whom were mothers and of low income. Adolescents who said they took part in many American behaviors and few Hispanic behaviors had the poorest family functioning in ninth grade. Most parents engaged in many Hispanic behaviors and few American behaviors. The discrepancy in reports of family functioning between adolescents and parents predicted adolescent cigarette smoking, sexual activity, and unprotected sex. The findings have implications for intervention efforts.

New Mothers' Sensitivity

How sensitive a new mother is to her baby's needs is critically important for both the baby's well-being and the new relationship the mother is building with her baby. From birth, the quality of the mother–baby relationship, or attachment, is thought to provide the foundation for the baby's emotional and relationship health throughout life. **Ablow, Marks, Feldman, and Huffman** (p. 1373) looked at more than 50 primarily European-American, middle-class women in the last trimester of their first pregnancies and when their babies were 9 months. They find that expectant women's state of mind in terms of attachment is associated with their physiological reactions to the sounds of crying babies: Soon-to-be moms with dismissive states of mind responded to crying babies in a way that suggests withdrawal or avoidant instincts, while women with autonomous states of mind tended to respond to the cries with a physiological pattern of approach instincts. The findings can inform intervention efforts to help new moms.

Visual Processing

Young infants face the challenge of learning in a complex and confusing world. In two experiments with about 40 infants from predominantly European, middle-class families, **Addyman and Mareschal** (p. 1137) show that 5-month-olds are sensitive to local redundancy in visual–temporal sequences—the lack of predictability in a sequence of stimuli that kept changing kept them looking. In demonstrating that babies are spontaneously aware of complexity and randomness in visual sequences, the study suggests that infants have basic attentional mechanisms that guide their learning.

Research has shown that individual differences in looking behavior can be identified in infancy and that these differences may be associated with efficiency of information processing. Infants who display brief visual fixations are labeled short lookers and are more likely to recognize a familiar stimulus than infants who display long visual fixations, labeled long lookers. In their study of about 20 primarily White, middle-class 6-month-olds, **Guy, Reynolds, and Zhang** (p. 1392) investigated the possibility that short and long lookers use different processing strategies and examined via electroencephalogram changes in brain activity associated with attention and memory. They find that short lookers attend to the global properties (overall configuration) of a stimulus, while long lookers attend to the local elements (smaller features) that compose the stimulus. This study replicates previous literature, but does so with brain-based measurement rather than behavioral data.

Being able to decode people's facial expressions is critical for social interaction. Between ages 5 and

12, significant changes occur in how the brain processes faces—for example, children become more accurate at identifying facial expressions of emotion—yet we don't know a lot about the mechanisms underlying these developmental changes. **Birmingham et al.** (p. 1407) used a new approach to examine the strategies used by about 130 Canadian children (ages 5–12) and young adults to selectively attend to different parts of the face and how these strategies change across development. Findings: At ages 11–12, children begin to look less often at the mouth and more at the eyes to understand others' mental states, and as they grow older, they look more at the left eye than the right, which allows them to make more sophisticated judgments about how facial expressions mirror emotions.

Sibling Dynamics

The social understanding known as theory of mind develops earlier in children with at least one sibling than in only children. That's the conclusion of **McAlister and Peterson** (p. 1442), who also find that having two or more siblings is even better. Their longitudinal study (of about 160 Australian ethnic-majority 3- to 5-year-olds from a variety of economic backgrounds) also finds that children with more siblings master executive function skills better than those with fewer or none. However, according to their research, while having siblings seems to positively influence the development of theory of mind, in the case of executive function, later developmental gains seem to be channeled at least partially through earlier theory of mind development. The results highlight the value of including assessment of sibling status in studies looking at the relation between theory of mind and executive function.

Children's experiences of causing harm to others may serve as key contexts for moral development. **Recchia, Wainryb, and Pasupathi** (p. 1459) examined about 100 middle-income children's and adolescents' open-ended accounts of harming their younger siblings and friends (children were 7, 11, and 16 and primarily White). They find that children's experiences harming their friends are distinct from their experiences harming siblings. Children described transgressions against friends as more unusual, unforeseeable, and circumstantial. Harm against siblings was said to occur when siblings engaged in obviously offensive behavior that was typical of the relationship. By contrasting children's accounts of hurting or upsetting their friends and siblings across a broad age range, the study provides insights into how youngsters' experiences with individuals of different ages contribute to their understanding of themselves, others, and relationships.

Summing Up

It's not how smart students are but how motivated they are and how they study that determines their *growth* in math achievement. That's what **Murayama, Pekrun, Lichtenfeld, and vom Hofe** (p. 1475) find in their examination of six annual waves of data from a German longitudinal study assessing math ability in 3,520 students in grades 5–10. Intelligence was strongly linked to students' math achievement, but only in the initial development of competence in the subject. Motivation and study skills turned out to be more important factors in terms of students' growth in math. Students who felt competent; were intrinsically motivated; used skills like summarizing, explaining, and making connections to other materials; and avoided rote learning showed more growth in math achievement than those who didn't. In contrast, after taking into account demographic characteristics, students' intelligence had no relation to growth in math achievement.

Anne Bridgman

Child Development, July/August 2013, Volume 84, Number 4, Pages 1120–1136

Social Equity Theory and Racial-Ethnic Achievement Gaps

Clark McKown
Rush University Medical Center

In the United States, racial-ethnic differences on tests of school readiness and academic achievement continue. A complete understanding of the origins of racial-ethnic achievement gaps is still lacking. This article describes social equity theory (SET), which proposes that racial-ethnic achievement gaps originate from two kinds of social process, direct and signal influences, that these two kinds of processes operate across developmental contexts, and that the kind of influence and the setting in which they are enacted change with age. Evidence supporting each of SET's key propositions is discussed in the context of a critical review of research on the Black–White achievement gap. Specific developmental hypotheses derived from SET are described, along with proposed standards of evidence for testing those hypotheses.

This article offers an account of the varied social processes that contribute to mean differences in test scores between children from different racial-ethnic groups, spanning preschool to high school (Jencks & Phillips, 1998; Lee, 2002; Lee & Burkam, 2002). In the United States, on measures of school readiness and academic achievement, Asian Americans achieve higher average scores than White students, who in turn achieve higher average scores than their Black and Latino peers (Jencks & Phillips, 1998; Lee, 2002; Reardon & Galindo, 2009). Because it has received a great deal of attention, the Black–White achievement gap is used to illustrate the major points of this article. The broad goal of this article, however, is to propose a model that applies to a variety of racial-ethnic and other achievement gaps.

Racial-ethnic achievement gaps are substantial life-span phenomena. By all accounts, the magnitude of the Black–White achievement gap is considerable, ranging from .5 to 1.0 *SD*, depending on the sample and the measure (Jencks & Phillips, 1998; Lee, 2002; Reardon & Galindo, 2009; Reardon & Robinson, 2007; Vanneman, Hamilton, Anderson, & Rahman, 2009). The Black–White achievement gap affects individuals and the generation to which they belong, beginning in early childhood and spanning all educational levels (Farkas, 2003). In terms of school readiness, research has consistently demonstrated that prior to school entry, Black students achieve lower average scores than White students (Brooks-Gunn, Klebanov, & Duncan, 1996; Duncan, Brooks-Gunn, & Klebanov, 1994; Lee & Burkam, 2002). Over the course of elementary school, the Black–White achievement gap appears to grow (Farkas, 2003; Fryer & Levitt, 2004, 2005; Phillips, Crouse, & Ralph, 1998).

The Black–White achievement gap is a highly consequential social problem. School readiness and academic achievement are associated with the kinds of jobs and wages people are able to secure. Racial-ethnic achievement gaps that begin at school entry and persist through school completion thus can influence racial-ethnic gaps in socioeconomic status (SES) across the life span (e.g., Levin, 2009; Reardon & Robinson, 2007). In turn, SES is robustly associated with health (Adler, Boyce, Chesney, & Cohen, 1994; Levin, 2009). Furthermore, the health of any democratic society is predicated on the ability of its population to make informed choices at the ballot box. When large segments of the population are inadequately educated, democracy's health is at risk.

As with any social problem, how policy makers, practitioners, and the public formulate the Black–White achievement gap's cause will guide what is done, and what is not done, to solve the problem (Humphreys & Rappaport, 1993; McKown, 2005;

This study was made possible by a William T. Grant Foundation Scholar's Award to Clark McKown. Thanks to Laura Gumbiner, Anne Gregory, Stephen Quintana, Michael Strambler, and Rhona Weinstein for their helpful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Clark McKown, Rush NeuroBehavioral Center, 4711 Golf Road, Suite 1100, Skokie, IL 60076. Electronic mail may be sent to Clark_A_McKown@rush.edu.

© 2012 The Author
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0002
DOI: 10.1111/cdev.12033

Seidman, 1983). Almost all accounts of racial-ethnic achievement gaps acknowledge the influence of multiple processes in multiple settings (e.g., Boykin, 1986; Garcia-Coll, 1990; Weinstein, 2002; Weinstein & McKown, 1998). Nevertheless, much prior work on the Black–White achievement gap has focused on specific processes operating in a limited range of contexts. For example, some have argued that racial differences in genetic endowment cause the gap (Jensen, 1969; Rushton & Jensen, 2005). Others have focused on the contribution of SES and family factors (Brooks-Gunn, Klebanov, Smith, Duncan, & Lee, 2003), cultural values (Ogbu, 2002; Thernstrom & Thernstrom, 2002), academic stereotypes (Steele & Aronson, 1995), and degree of match between home and school environments (Brice-Heath, 1983; Tharp, 1989).

More focused accounts are sometimes criticized because they do not account for the entire racial-ethnic achievement gap, even though they are often interpreted as doing so. For example, in a critical review, Sackett, Hardison, and Cullen (2004) acknowledged that one social process—stereotype threat—can depress Black students' test scores. However, they strongly questioned common interpretations of the research that stereotype threat is the sole cause of the Black–White test score gap, and that if it were eliminated, the gap would therefore disappear.

That theory and research have developed in a targeted way is understandable—evaluating multifactorial accounts of the gap's origin is challenging. The conceptual challenge is to broaden inquiry without mounting atheoretical "fishing expeditions" that provide few insights. The practical challenge is to muster resources to meaningfully evaluate the combined influences of multiple social processes across key developmental contexts. What is needed is a theory broad and flexible enough to account for varied sources of social influence, yet specific enough to provide a parsimonious account of the achievement gap.

The purpose of this article is to describe social equity theory (SET), a novel account of social processes that together give rise to racial-ethnic achievement gaps. SET provides a heuristic for organizing and integrating research on racial-ethnic achievement gaps. SET is proposed as a theory because: (a) it makes specific commitments about the social processes that are relevant to understanding the achievement gap, (b) it offers a causal explanatory framework to explain the racial-ethnic achievement gap, and (c) it generates specific, falsifiable hypotheses.

SET Defined

SET describes social processes that contribute to racial-ethnic achievement gaps. SET explains mean differences in achievement by members of different racial-ethnic groups. There is, of course, considerable within-group variability in academic readiness and achievement. Within-group and between-group differences in academic readiness and achievement are distinct. SET does not address the causes of individual variability in achievement. Rather, it explains factors that together create between-group differences in school readiness and achievement.

In the context of SET, the term *social processes* refers to transactions between individuals, including verbal and nonverbal communication directed from one person to another. The term *social processes* also refers to communications between individuals and social settings, in which an event or characteristic of the setting—apart from interpersonal interactions—communicates something of social consequence. For example, a poster depicting a civil rights event on a classroom wall may communicate something of social consequence without involving an *interpersonal* communicative transaction.

SET adopts an ecological perspective (Bronfenbrenner, 1977; Kelly, 1987; Spencer, 1999; Weinstein, 2002) to formulate what social processes in what contexts create and maintain racial-ethnic achievement gaps. SET includes four propositions about the origins of racial-ethnic achievement gaps:

1. Two classes of social process influence racial-ethnic achievement gaps. *Direct influences* are social processes that support achievement. Direct influences contribute to the racial-ethnic achievement gap when they are distributed differently to people from different racial-ethnic groups. *Signal influences* are cues that communicate negative expectations about a child's racial-ethnic group. When children from negatively stereotyped groups detect such cues, this can erode achievement.
2. Signal influences depend on children's ability to detect cues signaling a stereotyped expectation, and this ability increases significantly during the elementary grades.
3. Social processes affecting the achievement gap operate across a limited range of key developmental settings, and the relevant settings change lawfully with age.
4. Together, relevant direct and signal influences across developmental contexts account for the achievement gap.

From these propositions, it is possible to develop specific and falsifiable developmental hypotheses about the origins of racial-ethnic achievement gaps. What follows is a review of evidence supporting the existence of direct and signal influences on the achievement gap, their combined influence, and the contexts in which they influence racial-ethnic achievement gaps. The review focuses on research into the Black–White achievement gap. A subsequent section articulates specific hypotheses uniquely derived from SET. The article finishes with proposed standards of evidence to guide future work on SET and its component parts.

Evidence in Support of SET Propositions

Direct Influence on the Achievement Gap

Direct influences, defined. Direct influences are social processes that promote academic achievement similarly for all children in all racial-ethnic groups. SET proposes that direct influences contribute to racial-ethnic achievement gaps when they are systematically and unequally distributed to members of different racial-ethnic groups. Direct influences on racial-ethnic achievement gaps may unfold at home, in school, with peers, and in neighborhoods.

Home. In the home, particularly for young children, parent–child interactions and the relationships within which they take place are important contexts for children's development. A substantial body of research suggests that the characteristics and quality of parent–child interactions influences preacademic and academic outcomes. For example, Baumrind (1966; see also Baumrind & Black, 1967), focusing largely on White middle-class samples, examined the relation between parenting styles and a variety of life outcomes among children. They examined two dimensions of parenting style: supportiveness and demandingness. Supportiveness referred to the level of expressed parental love, nurturance, and responsiveness. Demandingness referred to the clarity of rules and the firmness and fairness of parental discipline practices. They found that when parents provided a combination of high support and high demand—what they called "authoritative" parenting—children had better average academic, social, and emotional outcomes.

There is some evidence that what normative parenting is varies from one racial-ethnic group to the next. Mandara and Murray (2002) found that Baumrind's (1966) parenting styles typology characterized Black parenting, but that Black authoritative parenting included greater demand and less compromise than White authoritative parenting. In addition, the relation between parenting and academic outcomes can differ for different groups. For example, at least two independent research groups have found that among Black children, the higher the overall neighborhood distress, the more strict parenting is associated with positive academic outcomes (Baldwin, Baldwin, & Cole, 1990; Gonzales, Cauce, Friedman, & Mason, 1996).

Prior research thus suggests that the same parenting practices do not always promote the same academic outcomes for children from different racial-ethnic groups, leaving open the question of whether parenting is a direct influence. The literature cited above suggests that what is *optimal* parenting for the development of academic outcomes may be different for children from different racial-ethnic groups and in different contexts. However, the same research suggests that for children from all racial-ethnic groups, achievement-supporting parenting—in whatever form that takes for the child's racial-ethnic group and context—plays a key role in the development of children's preacademic and academic skills.

For optimal parenting practices to exert a direct influence on the Black–White achievement gap, it must be present at different average levels in Black and White homes. In an important review article, Brooks-Gunn and Markman (2005) reported that there are racial-ethnic differences in parenting practices. They reported that, compared to White parents, Black parents on average engaged less frequent warm and sensitive parenting and more frequent negative and intrusive parenting. It is important to note that effect sizes were small, indicating a high degree of overlap in the distributions of parenting behavior. In addition, they reported differences in the frequency and richness of language exposure and availability of books and other stimulating materials, with Black parents on average engaging in less frequent conversation and having fewer books and other media available.

Parenting is thus associated with preacademic and academic skill development, and optimal parenting is, on average, more available to White children than to Black children. That does not mean that these factors necessarily explain the Black–White achievement gap. It is possible, for example, that parenting and achievement gaps both arise from a third variable, such as Black–White socioeconomic differences. If this were true, in empirical models with SES and parenting as independent variables and academic score as the dependent variable, SES, but not parenting, would be associated with academic outcomes. Existing evidence runs contrary to this conclusion. Examining the ECLS–K data, when Lee and Burkam (2002) controlled for SES, the magnitude of the racial-ethnic school readiness gap was reduced, but not eliminated. Similarly, in secondary analyses of several large data sets, Magnuson and Duncan (2006) found that SES accounts for some, but not all, of the Black–White test score gap among young children. Furthermore, in middle childhood and adolescence, accounting for SES reduces, but does not eliminate the Black–White test score gap (Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998).

When SES and racial-ethnic differences in parenting practices are accounted for, the size of the school readiness gap is dramatically reduced. For example, two studies found that among preschoolers, maternal warmth and engagement account for much of the Black–White gap that remained when SES was accounted for (Brooks-Gunn et al., 1996, 2003). Another study found that when controlling for SES, all the residual early childhood Black–White reading readiness gap and much of the residual math readiness gap was accounted for by family factors such as the frequency of reading (Lee & Burkam, 2002).

Parenting also helps explain the Black–White achievement gap in older children. For example, Mandara, Varner, Greene, and Richman (2009) examined the relation between parenting practices and the Black–White adolescent achievement gap. They examined multiple parenting practices, such as involving adolescents in decision making, parental monitoring, number of household chores, school orientation, and maternal warmth. They reported that: (a) White parents generally engage in more achievement-promoting practices than Black parents, (b) this difference is largely accounted for by racial-ethnic differences in SES, and (c) accounting for multiple parenting practices that promote achievement explains the entire adolescent achievement gap (Mandara et al., 2009).

In summary, optimal parenting practices: (a) promote preacademic and academic outcomes for all children; (b) are systematically and unequally distributed by racial-ethnic group; (c) are not an artifact of one critical potential confound, SES; and (d) account for much of the achievement gap that remains when SES is accounted for. Taken together, these findings suggest that parenting serves as a direct influence on the Black–White achievement gap.

School. The quality of instruction and the quality of student–teacher relationships may act as direct influences on racial-ethnic achievement gaps. Each of these school factors affects achievement similarly for students from all racial-ethnic groups. In addition, high-quality instruction and positive student–teacher relationships are more available for White than for Black students.

These direct influences can operate both between and within schools (Hanushek & Rivkin, 2009; Page, Murnane & Willett, 2008). In terms of between-school influences, for example, compared to White students, Black students attend schools in which instructional quality and teacher skill are, on average, lower (Clotfelter, Ladd, & Vigdor, 2004). Similarly, Lee and Burkam (2002) reported that at school entry, children from different racial-ethnic groups benefited from different levels of teacher experience, with African American children being assigned to less experienced teachers than their White and Asian peers. Thus, systematic racial-ethnic differences in the availability of experienced teachers and high-quality instruction between schools may exert a direct influence on the achievement gap.

Within-school processes may also exert a direct influence on the achievement gap. One important within-school process that can contribute to the achievement gap is mean Black–White differences in the quality of instruction. For example, throughout the primary and secondary grades, Black children are more often assigned to lower tracks than their White peers (Entwisle, Alexander, & Olson, 1997; Lucas & Berends, 2002). Children in different tracks are in turn exposed to different quality instruction, with students in higher tracks exposed to interesting and challenging material. Students in lower tracks, in contrast, are exposed to work that is repetitive, uninteresting, and focused on behavioral control (Boykin, 1986; Oakes, 2005; Weinstein, 2002). The quality of curriculum is thus a direct influence because it is associated with academic outcomes and it varies systematically by racial-ethnic group.

Black–White mean differences in the quality of relationships between teachers and students can also contribute to the achievement gap. The level of closeness and caring children and adolescents experience with their teachers influences academic outcomes (Gregory & Weinstein, 2004; Hamre & Pianta, 2005; Weinstein, 2002). On average, White students enjoy closer and more caring relationship with their teachers than Black students. For example, in a study that included 197 preschool and kindergarten teachers, Saft and Pianta (2001) found that White students enjoy less conflict with and dependency on their teachers than their Black peers. Similarly, in a study that included 607 academically at-risk first graders, teachers reported higher quality relationships with White compared to Black students (Hughes, Gleason, & Zhang, 2005). Others have found that teachers on average expect more of White students than Black students with similar records of achievement (Baron, Tom, & Cooper, 1985; Ferguson, 2008; McKown & Weinstein, 2008). Thus, there is evidence that the quality of teacher–student relationships may be a direct influence on the achievement gap: Those relationships are associated with academic outcomes and are, on average, higher quality for White students than their Black peers.

Educational field trials have provided evidence that equalizing instructional quality can reduce the gap. Success for All, a literacy curriculum that combines intensive focus on phonics and adherence to particular instructional practices, has been shown to reduce the achievement gap. Furthermore, the longer students are exposed to the curriculum, the greater the reduction in the gap (Slavin & Madden, 2001). Similarly, KIPP schools, a small network of charter schools that have extended hours, culturally appropriate instructional practices, individualized mentoring, and structured instruction, have shown promising evidence of promoting racial-ethnic minority achievement (Educational Policy Institute, 2005). These lines of evidence suggest that the quality of instruction is a direct influence on the achievement gap, with differential allocation of instructional quality contributing to Black–White differences in learning and achievement.

In summary, several school-based resources have a critical influence on achievement, including the quality of teachers, curriculum, and teacher–student relationships. There is evidence that White students, on average, benefit from better teachers, more challenging curriculum, and better relationships with teachers than their Black peers with similar records of achievement. Thus, these school factors likely exert a direct influence on the Black–White achievement gap.

Peers. Peer relationships and peer norms may exert a direct influence on the achievement gap. For example, ethnographic research suggests that for Black youth, the stigma associated with academic ambition and its connection with White culture can contribute to Black underachievement (Austen-Smith & Fryer, 2005; Fordham & Ogbu, 1986; Majors & Billson, 1993). Consistent with this formulation, Ferguson (2008) reported survey results suggesting that Black suburban high school students were more likely than their White peers to report withholding academic effort because of concern about how others might view them. Similarly, Graham (2001) reported that some racial-ethnic minority students value low-performing peers more than high-performing peers. Other studies, however, find less support for the effects of peer culture on the achievement gap (Ainsworth-Darnell & Downey, 1998; Cook & Ludwig, 1998). In contrast to the literatures on the influence of family factors and schooling, less empirical research has examined the influence of peers on the achievement gap. As a result, it remains largely an open question what peer influences most strongly influence achievement at what age, whether those influences are present in different amounts among children from different racial-ethnic groups, and overall, how much peer influences contribute to the achievement gap.

Neighborhood. Neighborhoods may exert a direct influence on the achievement gap. For example, the "collective efficacy" of neighborhoods, defined as the level of community cohesion and social support, may influence a variety of outcomes, including achievement (Sampson, Raudenbush, & Earls, 1997). In communities with strong cohesion, neighbors support one another as they work to achieve a range of goals, including the education of their children. Cook, Herman, Phillips, and Settersten (2002) found that the higher the neighborhood cohesion, the better students performed in school. Furthermore, children from different racial-ethnic groups remain largely segregated from one another within and across neighborhoods (Reardon et al., 2009) that differ in levels of achievement-promoting neighborhood cohesion. It is therefore plausible that neighborhood cohesion would exert a direct influence on the achievement gap. However, further research is necessary before a strong conclusion can be drawn.

Factors outside the realm of direct influences. Many phenomena partially fulfill the definition of a direct influence, but are not direct influences as defined by SET. Many of these factors are critical contexts that may themselves influence the proximal social transactions described by SET. For example, family structure (McLanahan & Percheski, 2008), SES (Yeung & Conley, 2008), and family wealth (Orr, 2003) are associated with achievement. In addition, there are mean racial-ethnic differences in family structure, SES, and family wealth. These phenomena thus embody two characteristics of direct influences: (a) they are associated with achievement, and (b) they systematically differ between racial-ethnic groups.

However, none of these phenomena is a *social process*, defined as a communication between individuals or between individuals and settings. As a result, they do not meet SET's definition of a direct influence. SET is thus narrowly tailored to explain the proximal social processes influencing the achievement gap. At the same time, there is very little doubt that broader contexts play a critical role in shaping the achievement gap. SET conceptualizes social processes as the mechanisms through which broader contexts shape the achievement gap. This conceptualization is consistent with work suggesting, for example, that the relation between SES and developmental outcomes is mediated by more proximal social processes (Brooks-Gunn & Duncan, 1997; Garcia-Coll, 1990; Guo, Brooks-Gunn, & Harris, 1996).

Summary. Social processes at home, at school, in the peer group, and in neighborhoods are associated with academic outcomes. Furthermore, these social processes are distributed unequally to children from different racial-ethnic groups. SET predicts that children from different racial-ethnic groups with similar exposure to these direct influences will achieve more similarly than children who have different exposure to those direct influences. However, SET also predicts that accounting for multiple direct influences will, in general, *reduce, but not eliminate racial-ethnic* differences in achievement. In addition to direct influences, other factors, called signal influences, contribute to the gap. As a result, even a comprehensive account of direct influences would not entirely explain the gap, particularly in late elementary school and beyond.

Signal Influences on the Achievement Gap

Signal influences, defined. Signal influences are social events that signal to members of negatively stereotyped groups that they are devalued because of their group membership (Inzlicht & Ben-Zeev, 2000; Mendoza-Denton, Downey, Purdie, Davis, & Pietrzak, 2002; Steele, 1997; Steele & Aronson, 1995). Signal influences may be overt, such as when a person is told directly that people from his or her group are not capable. Signal influences may also be ambiguous, such as when one person communicates through nonverbal behavior negative feelings toward another. Settings may also propagate signal influences outside of the context of an interpersonal encounter. For example, exposure to material that makes a child's racial-ethnic group membership salient can trigger signal influences with no interpersonal interaction (Ambady, Shih, Kim, & Pittinsky, 2001). When the target of an interpersonal or setting signal is from a stereotyped racial-ethnic group, the target may interpret that signal as a communication that the target is devalued because of his or her racial-ethnic group membership. Such a signal can interfere with performance in a stereotyped domain (Steele & Aronson, 1995). To someone from a nonstereotyped group, the same social events will not be interpreted as devaluing a self-relevant identity.

Signal influences are transactional. They do not exist solely in the mind of stereotyped individuals. Neither do they exist solely in the social setting with which the individual is engaged. Consistent with the definition of stereotype threat (Steele & Aronson, 1995), signal influences occur when an event in a social setting activates a concern in the mind of a stereotyped individual—consciously or not—that he or she may be judged on the basis of the stereotype. This process is distinct from self-stereotyping processes, which are presumed to occur in the mind of the stereotyped individual. It is also distinct from direct influences, although some social processes, described later, can operate as either direct or signal influences.

A paradigmatic example from the stereotype threat literature is that of the standardized testing situation. Although outwardly the same for all children, when testing conditions highlight that a challenging test is diagnostic of ability, members of academically stereotyped groups may become concerned that their performance will be judged in light of a stereotype about their group's intellectual ability, in turn lowering performance (McKown & Strambler, 2009; McKown & Weinstein, 2003; Steele & Aronson, 1995). The same event may communicate different meanings to children, depending on the prevailing cultural stereotype about each child's racial-ethnic group. For children from academically stereotyped racial-ethnic groups, the testing situation may signal the possibility of being devalued, which, in turn, can hamper performance.

Routine signal events. What characteristics of developmental contexts are likely to propagate signal influences on children's achievement, producing racial-ethnic differences in achievement? Experimental evidence has identified several kinds of events that propagate signal influences, including characterizing a test as diagnostic of natural ability (McKown & Strambler, 2009; McKown & Weinstein, 2003, Study 2; Steele & Aronson, 1995, Experiment 1), cuing participants to think about a stereotyped self-relevant identity before completing a task (Ambady et al., 2001), and telling participants directly that members of their group routinely perform worse on a task than members of other groups (Aronson et al., 1999). More commonplace factors, such as group composition, can create signal influences on academic performance, with minority status itself potentially signaling stereotyped expectations to minority group members (Inzlicht & Ben-Zeev, 2000; Murphy, Steele, & Gross, 2007).

Signal influences have been consistently demonstrated in the lab. It is less clear what events in daily life communicate to children that their intellectual ability is devalued because of their racial-ethnic group. Research on interpersonal expectancy effects suggests that teacher expectations may exert signal influences. In many classrooms, teachers expect more of White and Asian children than Black and Latino children with similar records of achievement (Baron et al., 1985; McKown & Weinstein, 2008; Rubovits & Maehr, 1973). Teacher differential treatment toward children from different racial-ethnic groups can be interpreted by children as a signal that the intellectual ability of their group is devalued. If this were so, the relation between teacher expectations and student achievement would be stronger for children from stereotyped groups than for their nonstereotyped peers. Indeed, the relation between teacher expectations and student achievement is more robust among Black than White students (Jussim, Eccles, & Madon, 1996), with low teacher expectations associated more strongly with negative academic outcomes for Black than White students (McKown & Weinstein, 2002).

It is conceivable that many events and interactions in the daily routine of schools signal to students from stereotyped groups that their intellectual ability is devalued. Those events may affect the extent to which students who are the targets of stereotypes feel that school is fair. Indeed, correlational studies suggest that high levels of perceived unfairness at school are related to higher rates of disciplinary referrals and disruptive behavior, and this effect is particularly strong for Black and Latino students (Gottfredson, 2001; Gregory & Weinstein, 2008; Tyler & Huo, 2000). When children from academically stereotyped racial-ethnic groups believe teachers are unfair, this belief may activate cultural narratives about racial injustice, signaling that they are devalued because of their ethnicity. This may in turn have a negative impact on the academic achievement of children from stereotyped racial-ethnic groups.

Signal influences and interpretive skill. Signal influences specifically and exclusively affect children from negatively stereotyped groups. Negative stereotypes about the academic ability of Blacks persist (Bobo, 2001), although it is no longer socially acceptable to express those stereotypes directly (Dovidio & Gaertner, 1998). Furthermore, many people who hold overtly egalitarian beliefs nonetheless may hold implicit stereotypes that lead them to appraise and respond to people on the basis of their race or ethnicity (Baron & Banaji, 2006; Gaertner & Dovidio, 1986; McConahay, Hardee, & Batts, 1981). Social norms against racism and the persistence of implicit stereotypes mean that in the contemporary American context, stereotypes are often expressed ambiguously (Gaertner & Dovidio, 1986). For example, a White teacher who is formal and nervous when interacting with a particular Black student may be expressing a formal, socially anxious style. Alternatively, she may feel that Black students are not capable, and her formality and discomfort are expressions of this stereotypic belief (McKown & Strambler, 2009). How the student interprets the origin of the teacher's behavior is critical in determining whether the teacher will propagate a signal influence.

The consequences of ambiguous signal events depend on children's awareness of stereotypes about their group. For ambiguous signals to propagate signal influences, children must know that others might devalue their ability because of their group membership. Equipped with this understanding, children can interpret ambiguous interpersonal behavior as a reflection of racial-ethnic stereotypes. Although these capacities must be present for ambiguous signals to have a negative influence on achievement, with adults and children, prior research suggests that stereotyped individuals do not have to be consciously aware of a signal influence for it to have a negative effect (Steele & Aronson, 1995). Rather, children and adults must be generally aware of self-relevant stereotypes in the performance domain for ambiguous signals to propagate negative performance (McKown & Strambler, 2009; McKown & Weinstein, 2003).

Children's thinking about others, about society, and about race and ethnicity changes dramatically in the first 12 years of life (Hirschfeld, 1996; Killen, Rutland, & Ruck, 2011; Killen & Stangor, 2001; Nesdale & Flesser, 2001; Quintana, 1998; Rutland, Cameron, Milne, & McGeorge, 2005). In turn, how children appraise and interpret social events affects their response to those events (e.g., McKown & Strambler, 2009; Spencer, 1999). By preschool, American children have developed a conception of race (Hirschfeld, 1996). By elementary school, children use group membership in their judgments about the fairness of social exclusion (Killen, Lee-Kim, McGlothlin, & Stangor, 2002). In this age range, children become aware of the relation between group membership and intergroup liking (Quintana, 1998). Specifically, between the ages of 6 and 10, children become aware that *others* endorse stereotypic beliefs that may devalue their group's intellectual ability (Brown & Bigler, 2005; Brown, Bigler, & Chu, 2010; McKown & Strambler, 2009; McKown & Weinstein, 2003; Quintana, 1998, 2008).

All these lines of evidence suggest that beginning around third grade, and commonly by the end of elementary school, typically developing children have the social knowledge and capacity to interpret events that, on their surface, are ambiguous in terms of whether they reflect biased beliefs or attitudes, as expressions of stereotypes or prejudices. The ability to interpret events as signals about what others believe about a child's group is, in turn, highly consequential.

For example, children's growing stereotype consciousness can affect their response to social and academic events. McKown and Strambler (2008) found that when children were aware of broadly held stereotypes, they were more likely to interpret negative interethnic encounters as reflecting racial animus. Furthermore, children from academically negatively stereotyped racial-ethnic groups (African American and Latino) *who were aware of broadly held cultural stereotypes* performed more poorly on a cognitive task when the task was described as diagnostic of ability than when the same task was described as a problem-solving task. For children from nonstereotyped racial-ethnic groups and for children from stereotyped racial-ethnic groups who were unaware of broadly held stereotypes, the conditions of testing did not affect performance (McKown & Strambler, 2009; McKown & Weinstein, 2003). These findings suggest that when children from stereotyped groups become aware of broadly held stereotypes, signal influences can hamper performance on cognitive and academic tasks.

Randomized field trials of interventions suggest that children and youth can be inoculated against naturally occurring signal influences, further demonstrating their existence and impact. For example, drawing from on motivation theory (Dweck, 1988), Good, Aronson, and Inzlicht (2003) found that teaching middle school girls to believe that intelligence depends on effort and that academic difficulties are situational reduced the effects of stereotype threat on math test performance. Drawing from self-affirmation theory (Steele, 1988), Cohen, Garcia, Apfel, and Master (2006) found when African American students participated in an exercise in which they wrote a paragraph about their values and why those values were important, they showed lower levels of stereotype activation and received higher grades than their African American peers in a control condition. By successfully inoculating children against signal influences outside the lab, these studies indirectly demonstrate the existence of signal influences in naturalistic settings.

Child Factors

The foregoing discussion of direct and signal influences emphasizes social processes and the social settings in which they take place. This raises the question of what role child factors play in causing and maintaining racial-ethnic achievement gaps. There is little doubt that children's individual biological and cognitive characteristics affect their academic outcomes. For example, the better children perform on cognitive tasks such as those represented in IQ tests, the better children do academically (Neisser et al., 1995). Furthermore, how children understand and process information, and therefore what they can learn, changes normatively with age (Piaget, 1972).

It is beyond the scope of this article to review the many ways individual child characteristics and developmental processes contribute to achievement. An underlying premise of SET is that mean racial-ethnic group differences in achievement are the product of social processes that transpire between children and peers, between children and adults, and between children and social settings. Group differences in IQ test scores are also presumed to arise from these same social forces. A corollary premise of SET is that although individual child factors influence within-group variability in achievement, they do not, in general, explain mean racial-ethnic group differences in achievement. There is one important exception: The impact of signal influences depends on children's awareness of broadly held societal stereotypes, which changes predictably with age (McKown & Strambler, 2008; McKown & Weinstein, 2003).

Overlap and Boundaries Between Direct and Signal Influences

Direct and signal influences are distinct mechanisms by which a social event may affect group differences on the Black–White achievement gap. Some events are clearly and only signal events. For example, the conditions of testing may signal to members of academically stereotyped racial-ethnic groups that their intellectual ability is devalued. Those conditions create unequal outcomes exclusively through their signal properties. Other events are clearly and only direct influences. For example, enrollment in schools that deliver high-quality instruction is an important resource for the development of talent (Hamre & Pianta, 2005; Weinstein, 2002). Higher quality instruction is associated with greater opportunities to learn and achieve. To the extent that children from different racial-ethnic groups are enrolled in different schools with instruction whose quality is systematically different, this will exert an influence on racial-ethnic achievement gaps through a direct influence on learning and achievement. If students are unaware that the allocation of resources is unequal, this form of discrimination will exert only a direct influence on racial-ethnic achievement gaps.

Some social phenomena can contribute to the achievement gap through direct influence, signal influence, or both (McKown, Gregory, & Weinstein, 2010). Research on teacher expectations provides an illustration. Teachers who have higher expectations of students tend to behave more warmly, provide better instruction, and offer greater student autonomy, all of which is associated with greater learning (Brophy & Good, 1970; Weinstein, 2002). In some classrooms, teachers expect more of their White students than they do of their Black students with similar records of achievement (Baron et al., 1985; McKown & Weinstein, 2008). Teacher expectations, allocated in part on the basis of child race, may lead teachers to provide better quality instruction to White students than to Black students with similar records of achievement. This may in turn lead to Black–White differences in achievement. In this case, the differential allocation of instructional quality is a direct influence, with no requirement that children notice the differential treatment for it to contribute to the achievement gap. However, it is also possible that some students, noticing the teacher's differential treatment of children from different racial-ethnic groups, will interpret the teacher's low expectations as reflecting the teacher's belief that some racial-ethnic groups are not capable. If so, this may lead to a signal influence. Teacher expectations may thus exert direct *and* signal influences on the Black–White achievement gap.

Although some experiences can operate as both direct and signal events, the distinction between the two kinds of influences is nonetheless useful. From a theoretical perspective, distinguishing direct and signal influences provides a heuristic for evaluating two distinct pathways through which events may influence the achievement gap. From an applied perspective, how we understand the mechanisms through which events affect the achievement gap will shape the policies and practices that are deployed to reduce the gap. For example, if teacher expectations are seen as exclusively exerting a direct influence on the gap, then the best prescription for reducing its influence is to equalize teacher behavior that communicates expectations. Attempts to do so have met with mixed results (Gottfredson, Marciniak, Birdseye, & Gottfredson, 1995; Weinstein et al., 1991). In contrast, if teacher expectations are seen as exerting a signal influence, then a prescription might be to inoculate students against the untoward effects of signal influences by, for example, teaching children that achievement is the product of hard work (Good et al., 2003), or having children affirm important positive self-relevant values (Cohen et al., 2006). If, as the evidence suggests, teacher expectations can exert direct and signal influences on the Black–White achievement gap, the focus might include equalizing teacher behavior *and* inoculating students.

Combined Effects of Direct and Signal Influences

SET proposes that multiple social processes operate together within and across key developmental contexts to influence racial-ethnic achievement gaps. Thus, with the goal of identifying the universe of influences on the achievement gap, it is insufficient to account only for direct influences or only for signal influences. It is also generally insufficient to account for direct and signal influences in a single context. SET incorporates the proposition that direct and signal influences can operate in any key developmental context. The specific processes may be similar or different in different contexts. In addition, the proportion of the variance in a group difference resulting from direct and signal influences may be different in each context and in each developmental phase. However, the theory proposes that only by identifying salient direct and signal influences in key developmental contexts and by combining effects across contexts can we come to a full accounting of any racial-ethnic achievement gap.

Each of SET's propositions is supported from the empirical literature. However, the overall proposition that the achievement gap is better accounted for by combining the effects of different social processes occurring in different key developmental contexts remains largely unexamined. Some researchers have examined the combined association of more than one direct influence on racial-ethnic achievement gaps (e.g., Brooks-Gunn et al., 1996, 2003; Lee & Burkam, 2002). However, no studies have examined multiple signal influences, or the combined effects of direct and signal influences across contexts. SET lends itself to the study of multiple influences across multiple contexts simultaneously. It is conceivable that influences within and across developmental contexts would combine additively or interactively to produce the Black–White achievement gap. How varied forces described by SET combine to produce the gap is an important area for future investigation.

Universe of Influential Developmental Contexts

A key proposition of SET is that social processes that cause and maintain the achievement gap may operate in and across multiple developmental contexts (Bronfenbrenner, 1977; Kelly, 1987; Weinstein, 2002). A context is defined as any setting in which an individual transacts social interactions with important others or in which setting characteristics signal information of social value. Developmental contexts in childhood include family, school, peer network, and neighborhood. During different developmental epochs, the contexts in which social processes influence the achievement gap will change. As children transition to adulthood, new contexts become important, including the workplace and marriage. Within each of these contexts are varied "microsettings," in any one of which, direct or signal influences may unfold. What contexts are the most important focus of inquiry and intervention should be guided by an understanding of the phenomenon of interest, theory, and the weight of evidence.

The earlier review of direct and signal influences suggests that there is good reason to focus on family, school, peers, and neighborhoods. Specifically, the literature suggests that: (a) family factors such as parenting practices influence the achievement gap; (b) some of the achievement gap can be attributed to racial-ethnic differences in the availability of high-quality instruction; (c) signal influences in naturalistic settings, particularly school, contribute to the achievement gap; (d) peers may exert an influence on children's academic achievement; and (e) neighborhood psychological resources influence achievement and may influence the achievement gap. Furthermore, the foregoing literature review suggests that direct and signal influences may unfold in any one of these developmental settings. One of the most substantial unanswered questions raised by SET is what direct and signal influences operate in what developmental settings at what age to explain the achievement gap.

Developmental Hypotheses Uniquely Derived From SET

SET leads to specific predictions about: (a) the relative influence of direct versus signal influences on the achievement gap at different ages, (b) what developmental contexts are likely to have the greatest influence on the achievement gap at what age, and (c) what set of direct and signal influences in what contexts are sufficient to account for the achievement gap.

Age-Related Differences in the Contribution of Direct and Signal Influences

SET predicts that the relative impact of direct and signal influences will change with age. In preschool through second grade, most children are not yet aware that others endorse stereotypic beliefs (McKown & Strambler, 2009; McKown & Weinstein, 2003; Quintana, 1998, 2008). In those early years, then, only the most overt expression of stereotypes could be interpreted by a child as being self-relevant. Under these developmental constraints, children cannot interpret more ambiguous cues that a child may be judged on the basis of group membership as a devaluation of that group. By third grade, most children have become aware of broadly held cultural stereotypes. Becoming aware of stereotypes is in turn associated with susceptibility to stereotype threat effects induced through test instructions (McKown & Strambler, 2009; McKown & Weinstein, 2003).

SET thus predicts that in preschool through second grade, racial-ethnic differences in test scores are equal to the combination, across key developmental contexts, of direct influences only. Why? Signal influences cannot affect children when they have not yet developed the capacity to detect those signals. As children develop the ability to interpret social cues as reflecting stereotypes, a full explanation of the achievement gap requires accounting for both direct and signal influences. Between third grade and adolescence, most children become aware of broadly held stereotypes. SET thus predicts that the relative impact of direct and signal influences on the Black–White achievement gap will shift from exclusively direct influences through second grade, to a preponderance of direct influences and some signal influences between third and fifth grades, to a more equal balance between direct and signal influences thereafter (see Figure 1).

The empirical literature is generally consistent with this hypothesis. Among preschoolers, accounting for multiple direct influences eliminates or nearly eliminates school readiness and test performance gaps (Brooks-Gunn et al., 2003; Lee & Burkam, 2002). In contrast, with middle school and high school students, controlling for direct influences is insufficient to account entirely for the achievement gap (Duncan & Magnuson, 2005; but see Mandara et al., 2009; Yeung & Pfeiffer, 2009). According to SET, this is because beyond second grade, controlling for direct influences on the Black–White achievement gap does not account for signal influences and thus leaves a residual portion of the gap unexplained. When children develop awareness of broadly held cultural stereotypes, to account for the entire gap requires accounting for both direct and signal influences (McKown & Strambler, 2009; McKown & Weinstein, 2003).

Age-Related Differences in the Contribution of Social Influences Across Contexts

Prior work suggests what settings are most influential in each epoch of a child's life. In infancy and toddlerhood, children spend most of their time and relational energy with parents and other caregivers. After the transition to elementary school, first teachers, then peers, play increasingly prominent roles in children's daily social experiences. Late in elementary school, as they become more independent, children may have more exposure to neighborhood influences. In adolescence, close friendships, peer networks, and romantic partners may become prominent influences, and parents exert less direct influence over their children's daily lives.

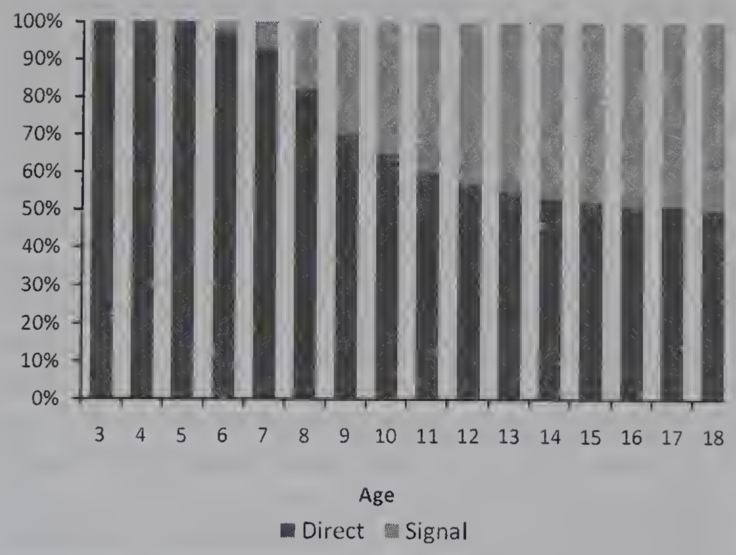


Figure 1. Hypothesized relation between age and the proportion of the achievement gap accounted for by direct and signal influences.

In light of these developmental changes, SET hypothesizes that prior to school entry, only home influences affect the Black–White achievement gap. After school entry and through second grade, social processes at home and at school affect the Black–White achievement gap, with home processes initially contributing more to the gap than school processes. As school becomes more academically focused through fifth grade, social processes in school play an increasing role in contributing to the gap. After second grade, peer relationships move from more transient play to more enduring "chumships." At this point, peers may begin to contribute to the gap, and this influence will increase in adolescence. In late elementary school, as children become more independent in the community, neighborhood effects on the achievement gap come into play, and grow incrementally through adolescence. This hypothesis is depicted in Figure 2.

Intervention Implications

Developmental hypotheses derived from SET provide the basis for creating a developmentally informed sequence of interventions. Specifically, to reduce the school readiness gap, preschool interventions should promote positive parenting and achievement-supporting home settings. From kindergarten through second grade, in addition to

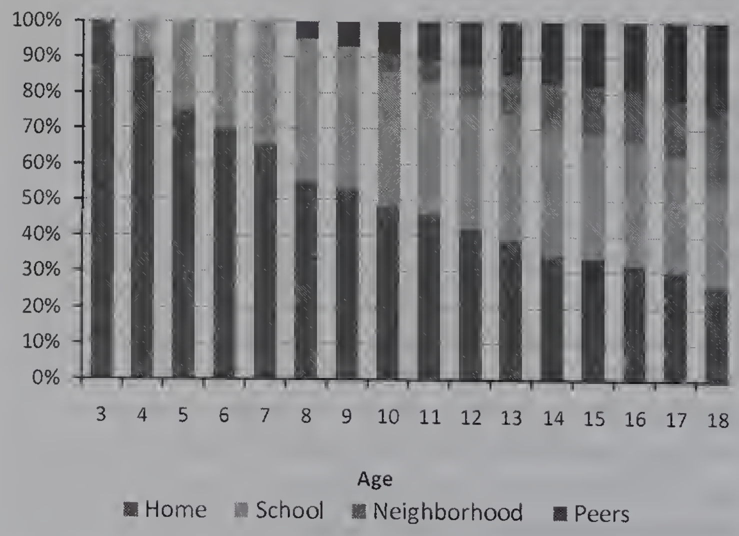


Figure 2. Hypothesized proportion of the achievement gap influenced in different developmental contexts, by age.

family support, interventions should focus on providing strong instruction and positive teacher–child relationships. From third grade on, in addition to parent support and strong academic support, interventions should inoculate children against the effects of signal influences and should endeavor to reduce the overall level of stereotyping at school. From third grade on, interventions should also be implemented to promote positive peer norms for achievement. From fifth grade on, in addition to the aforementioned kinds of support, interventions should increase neighborhood collective efficacy or, if that is not feasible, help children cope with neighborhood influences. SET predicts that if direct and signal influences salient in a particular setting and a developmental epoch are addressed, the Black–White achievement gap would be eliminated.

The Sufficiency of SET

Research on the achievement gap has been marked by a consistent partial accounting for the gap. SET holds promise to be a sufficient explanation for the Black–White achievement gap across developmental epochs. Not all possible developmental processes and theories are incorporated into SET. Furthermore, other social forces not incorporated into SET may contribute to racial-ethnic achievement gaps. As a result, two forms of SET are possible. If developmental phenomena and the full universe of social forces are sufficiently accounted for by SET, a strong form of the theory will withstand empirical scrutiny: Direct and signal influences, combined across key developmental contexts, will explain the entire Black–White achievement gap. If on the other hand, relevant developmental processes or social forces are not integrated into SET, a weak form of the theory may be more reasonable: Direct and signal influences, combined across key developmental contexts, will explain *more* of the achievement gap than current models do, but other forces, not accounted for by SET also contribute to the gap. Whether the strong or weak form of SET is more justified is a matter of empirical inquiry.

Standards of Evidence

Existing evidence supports many of the elements of SET. However, further research is needed to evaluate the theory as a whole. A judicious combination of naturalistic and experimental methods can be used to test the components of SET and the theory in its entirety. Three standards are required, one for each of the following: (a) evidence of direct influences, (b) evidence of signal influences, and (c) evidence supporting the whole of SET.

Evidence of Direct Influences

The standard of evidence for the presence of a direct influence on the Black–White achievement gap is as follows: If a social process has a direct influence, Black and White children who share similar exposure to the direct influence should achieve more similarly than children with different exposure to the direct influence. Furthermore, the relation between hypothesized direct influences and achievement should be similar for children from all racial-ethnic groups. For example, Black and White children who live in comparably cohesive neighborhoods should achieve more similarly than children who live in neighborhoods that differ in their cohesiveness. In statistical terms, controlling for direct influences should reduce the magnitude of the relation between racial-ethnic group membership and achievement, and there should be no ethnicity by direct influence interaction.

Naturalistic research examining the relation between direct influences and the Black–White achievement gap generally examines whether the magnitude of the relation between child race-ethnicity and achievement is reduced when hypothesized direct influences are included in a statistical model. For example, this method has been used to demonstrate that comparing children with similar home environments accounts for much of the Black–White achievement gap (Brooks-Gunn et al., 1996; Lee & Burkam, 2002). Further research using naturalistic methods will help clarify the nature and magnitude of the combined effect of direct influences on the gap. The benefit of naturalistic studies is that they tend to be more feasible than true experiments. The drawback is that they do not permit strong causal inferences.

True experiments can lead to inferences about the causal impact of direct influences. For example, the New Hope Project, a randomized antipoverty program, found that income supplements, child-care assistance, and health-care benefits to full-time working poor families led to academic, social, and emotional benefits to children, particularly boys (Huston et al., 2001, 2005). Experiments permit stronger inferences about the causal relation between hypothesized direct influences and achievement. Although they are well suited to isolating the causal influence of a single independent variable, true experiments can be costly.

Evidence of Signal Influences

If a setting characteristic has a signal influence on the achievement gap, it should be true that in the presence of the signal, members of stereotyped racial-ethnic groups only will perform more poorly than when the signal is absent. In statistical terms, signal influences will produce an Ethnicity × Signal Strength interaction. More specifically, the stronger the signal, the more poorly children from stereotyped groups will perform; in contrast, children from nonstereotyped groups will perform similarly regardless of the level of the signal.

As with direct influences, signal influences can be assessed using experimental or naturalistic designs. With regard to experimental studies, a raft of studies on stereotype threat has demonstrated that a variety of events can signal to children and adults from stereotyped groups that their intelligence is devalued, causing reduced performance (Nguyen & Ryan, 2008; Steele & Aronson, 1995; Walton & Spencer, 2009). Naturalistic methods could also be used to examine whether members of stereotyped groups, in the presence of presumed signal events, are more negatively affected than members of nonstereotyped groups. Experimental and naturalistic methods have been used, for example, in studies demonstrating that teachers expect more of White and Asian students than their Black and Latino peers with similar records of achievement (Baron et al., 1985; McKown & Weinstein, 2008) and that the relation between teacher expectations and achievement is stronger for Black students than for White students (Jussim et al., 1996), particularly when teacher expectations are negative (McKown & Weinstein, 2002).

Evidence Supporting SET as a Whole

Evidence supporting the constituent elements of SET does not necessarily support SET as a whole. Many studies presented in this article support the constituent elements of SET by demonstrating that a direct or signal influence accounts for *part of the Black–White achievement gap*. However, no study has totally accounted for the gap at all ages and no intervention or policy solution has totally eliminated it. A full assessment of SET will involve assessing the effect, across developmental contexts and ages, of multiple direct and signal influences. Ideally, such an examination would be undertaken using naturalistic designs *and* field experiments. Doing so would provide a fair test of the theory as a whole and would provide useful information about comprehensive strategies to reduce educational inequality.

Conclusion

SET is a general, flexible model of social processes that contribute to the American racial-ethnic achievement gaps. SET proposes that racial-ethnic achievement gaps are the result of the combination, across key developmental contexts, of direct and signal influences. A central goal in developing SET is to provide a parsimonious model from which hypotheses can be developed and systematically, programmatically, and empirically tested. Although the review of evidence presented in this article focused on the Black–White achievement gap, SET provides a framework for understanding, examining, and reducing any racial-ethnic achievement gap. Furthermore, SET may prove useful in explaining socioeconomic achievement gaps, and sex differences in science, math, and engineering occupations (Cheryan & Plaut, 2010). SET may also offer a useful heuristic for understanding the origins of inequalities in other domains, such as racial-ethnic differences in disciplinary referrals at school (Gregory, Skiba, & Noguera, 2010) and employment outcomes (Pager & Shepherd, 2008). The accuracy and general applicability of SET to other forms of inequality are empirical questions to be vigorously investigated.

Prior research on the Black–White achievement gap provides support for each of the constituent elements of SET. Skeptics will rightly conclude that support for SET's constituent elements is not the same as support for the theory as a whole. However, there have been no comprehensive assessments of SET as a singular explanation of the Black–White achievement gap. Although it poses methodological challenges, only by conducting such research can we understand the full range of social processes that contribute to this significant social problem. In so doing, we may come to more efficient and effective policies and practices to reduce inequalities that run afoul of core American ideals of fair play.

References

Adler, N. E., Boyce, T., Chesney, M. A., & Cohen, S. (1994). Socioeconomic status and heath: The challenge of the gradient. *American Psychologist*, *49*, 15–24.

Ainsworth-Darnell, J. W., & Downey, D. B. (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American Sociological Review*, *63*, 536–553.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*, 385–390.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46.

Austen-Smith, D., & Fryer, R. G. (2005). An economic analysis of "acting White." *Quarterly Journal of Economic Science*, *120*, 551–583.

Baldwin, A. L., Baldwin, C., & Cole, R. E. (1990). Stress-resistant families and stress-resistant children. In J. E. Rolf, A. S. Masten, D. Cicchetti, K. H. Nuechterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 257–280). New York: Cambridge University Press.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58.

Baron, R. M., Tom, D. Y., & Cooper, H. M. (1985). Social class, race, and teacher expectations. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 251–270). Hillsdale, NJ: Erlbaum.

Baumrind, D. (1966). Effects of authoritative control on child behavior. *Child Development*, *37*, 887–907.

Baumrind, D., & Black, A. (1967). Socialization practices associated with dimensions of competence in preschool boys and girls. *Child Development*, *38*, 291–327.

Bobo, L. (2001). Racial attitudes and relations at the close of the twentieth century. In N. J. Smelser, W. J. Wilson, & F. Mitchell (Eds.), *America becoming: Racial trends and heir consequences* (pp. 264–301). Washington, DC: National Academy Press.

Boykin, A. W. (1986). The triple quandary and the schooling of Afro-American children. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives* (pp. 57–92). Hillsdale, NJ: Erlbaum.

Brice-Heath, S. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: McGraw-Hill.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*, 513–531.

Brooks-Gunn, J., & Duncan, G. (1997). The effects of poverty on children. *The Future of Children*, *7*, 55–71.

Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development*, *67*, 396–408.

Brooks-Gunn, J., Klebanov, P. K., Smith, J., Duncan, G. J., & Lee, K. (2003). The Black-White test score gap in young children: Contributions of test and family characteristics. *Applied Developmental Science*, *7*, 239–252.

Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *The Future of Children*, *15*, 139–168.

Brophy, J., & Good, T. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, *61*, 365–374.

Brown, C. S., & Bigler, R. S. (2005). Children's perceptions of discrimination: A developmental model. *Child Development*, *76*, 533–553.

Brown, C. S., Bigler, R. S., & Chu, H. (2010). An experimental study of the correlates and consequences of perceiving oneself to be the target of gender discrimination. *Journal of Experimental Child Psychology*, *107*, 100–117.

Cheryan, S., & Plaut, V. C. (2010). Explaining underrepresentation: A theory of precluded interest. *Sex Roles*, *63*, 475–488.

Clotfelter, C., Ladd, H., & Vigdor, J. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? *Journal of Policy Analysis and Management*, *23*, 251–271.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, *313*, 1307–1310.

Cook, T., Herman, M. R., Phillips, M., & Settersten, R. A. (2002). Some ways in which neighborhoods, nuclear families, friendship groups, and schools jointly affect changes in adolescent development. *Child Development*, *73*, 1283–1309.

Cook, P. J., & Ludwig, J. (1998). The burden of "acting White": Do Black adolescents disparage academic achievement? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 375–400). Washington, DC: Brookings Institution.

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response*. Thousand Oaks, CA: Sage.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development*, *65*, 296–318.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, *15*, 35–54.

Dweck, C. S. (1988). Motivational processes affecting learning. *American Psychologist*, *41*, 1040–1048.

Educational Policy Institute. (2005). *Focus on results: An academic impact analysis of the Knowledge Is Power Program (KIPP)*. Virginia Beach, VA: Author.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Boulder, CO: Westview.

Farkas, F. (2003). Racial disparities and discrimination in education: What do we know, how do we know it, and what do we need to know? *Teachers College Record*, *6*, 1119–1146.

Ferguson, R. F. (2008). *Toward excellence with equity: An emerging vision for closing the achievement gap*. Cambridge, MA: Harvard Education Press.

Fordham, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the "burden of acting White." *Urban Review, 18*, 176–206.

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics, 86*, 447–464.

Fryer, R. G., & Levitt, S. D. (2005). *The Black-White test score gap through third grade* (NBER Working Paper No. 11049). Cambridge, MA: National Bureau of Economic Research.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). San Diego, CA: Academic Press.

Garcia-Coll, C. (1990). Developmental outcome of minority children: A process-oriented look into our beginnings. *Child Development, 61*, 270–289.

Gonzales, N. A., Cauce, A. M., Friedman, R. J., & Mason, C. A. (1996). Family, peer, and neighborhood influences among African-American adolescents: One-year prospective effects. *American Journal of Community Psychology, 24*, 365–387.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*, 645–662.

Gottfredson, D. C. (2001). *Schools and delinquency*. New York: Cambridge University Press.

Gottfredson, D. C., Marciniak, E., Birdseye, A. T., & Gottfredson, G. D. (1995). Increasing teacher expectations for student achievement. *Journal of Educational Research, 88*, 155–163.

Graham, S. (2001). Inferences about responsibility and values: Implication for academic motivation. In F. Salili & C. Y. Chiu (Eds.), *Student motivation: The culture and context of learning. Plenum series on human exceptionality* (pp. 31–59). Dordrecht, the Netherlands: Kluwer Academic.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39*, 59–68.

Gregory, A., & Weinstein, R. S. (2004). Connection and regulation at home and in school: Predicting growth in achievement for adolescents. *Journal of Adolescent Research, 19*, 405–427.

Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology, 46*, 455–475.

Guo, G., Brooks-Gunn, J., & Harris, K. M. (1996). Parents' labor force attachment and grade retention among urban black children. *Sociology of Education, 69*, 217–236.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949–967.

Hanushek, E. A., & Rivkin, S. G. (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of Policy Analysis and Management, 28*, 366–393.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.

Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of School Psychology, 43*, 303–320.

Humphreys, K., & Rappaport, J. (1993). From the community mental health movement to the war on drugs: A study in the definition of social problems. *American Psychologist, 48*, 892–901.

Huston, A. C., Duncan, G. J., Granger, R., Bos, J., McLoyd, V., Mistry, R., et al. (2001). Work-based antipoverty programs for parents can enhance the school performance and social behavior of children. *Child Development, 72*, 318–336.

Huston, A. C., Duncan, G. J., McLoyd, V. C., Crosby, D. A., Ripke, M. N., Weisner, T. S., et al. (2005). Impacts on children of a policy to promote employment and reduce poverty for low-income parents: New Hope after 5 years. *Developmental Psychology, 41*, 902–918.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365–371.

Jencks, C., & Phillips, M. (Eds.) (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.

Jensen, A. (1969). How much can we boost IQ achievement? *Harvard Educational Review, 39*, 1–123.

Jussim, L., Eccles, J. S., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology, 28*, 281–388.

Kelly, J. G. (1987). An ecological paradigm: Defining mental health consultation as a preventive service. *Prevention in Human Services, 4*, 1–36.

Killen, M., Lee-Kim, J., McGlothlin, H., & Stangor, C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the Society for Research in Child Development, 67*(4), 1–119.

Killen, M., Rutland, A., & Ruck, M. (2011). Promoting equity, tolerance, and justice in childhood. *SRCD Social Policy Report, 25*, 3–25.

Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development, 82*, 174–186.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31*, 3–12.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Levin, H. M. (2009). The economic payoff to investing in educational justice. *Educational Researcher, 38*, 5–20.

Lucas, S. R., & Berends, M. (2002). Sociodemographic diversity, correlated achievement, and de facto tracking. *Sociology of Education*, *75*, 328–348.

Magnuson, K., & Duncan, G. J. (2006). The role of family socioeconomic resources in the Black-White test score gap among young children. *Developmental Review*, *26*, 365–399.

Majors, R., & Billson, J. M. (1993). *Cool pose: The dilemmas of Black manhood in America*. New York: Touchstone Books/Simon & Schuster.

Mandara, J., & Murray, C. B. (2002). Development of an empirical typology of African American family functioning. *Journal of Family Psychology*, *16*, 318–337.

Mandara, J., Varner, F., Greene, N., & Richman, S. (2009). Intergenerational family predictors of the Black-White achievement gap. *Journal of Educational Psychology*, *101*, 867–878.

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, *25*, 563–579.

McKown, C. (2005). Applying ecological theory to advance the science and practice of school-based prejudice reduction interventions. *Educational Psychologist*, *40*, 177–189.

McKown, C., Gregory, A., & Weinstein, R. S. (2010). Expectations, stereotypes, and self-fulfilling prophecies in classroom and school life. In J. Meece & J. Eccles (Eds.), *Handbook of research on schools, schooling, and human development*. New York: Routledge.

McKown, C., & Strambler, M. J. (2008). Social influences on the ethnic achievement gap. In S. Quintana & C. McKown (Eds.), *Handbook of race, racism, and the developing child* (pp. 366–396). Hoboken, NJ: Wiley.

McKown, C., & Strambler, M. J. (2009). Developmental antecedents and social and academic consequences of stereotype-consciousness in middle childhood. *Child Development*, *80*, 1643–1659.

McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, *32*, 159–184.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype-consciousness in middle childhood. *Child Development*, *74*, 498–515.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, *46*, 235–261.

McLanahan, S., & Percheski, C. (2008). Family structure and the reproduction of inequality. *Annual Review of Sociology*, *34*, 257–276.

Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to status-based rejection: Implications for African American students' college experience. *Journal of Personality and Social Psychology*, *83*, 896–918.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879–885.

Neisser, U., Boodoo, G., Bouchard, T. K., Boykin, A.W., Brody, N., Ceci, S.J. (1995). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.

Nesdale, D., & Flesser, D. (2001). Social identify and the development of children's group attitudes. *Child Development*, *72*, 506–517.

Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314–1334.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.

Ogbu, J. (2002). Cultural amplifiers of intelligence: IQ and minority status in cross-cultural perspective. In J. M. Fish (Ed.), *Race and intelligence: Separating science from myth* (pp. 241–278). Mahwah, NJ: Erlbaum.

Orr, A. J. (2003). Black-White differences in achievement: The importance of wealth. *Sociology of Education*, *76*, 281–304.

Page, L. C., Murnane, R. J., & Willett, J. B. (2008). *Trends in the Black-White achievement gap: Clarifying the meaning of within and between-school achievement gaps* (NBER Working Paper No. 14213). Cambridge, MA: National Bureau of Economic Research.

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, *34*, 181–209.

Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 103–145). Washington, DC: Brookings Institution Press.

Phillips, M., Crouse, J., & Ralph, J. (1998). Does the Black-White test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 229–272). Washington, DC: Brookings Institution Press.

Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield Adams.

Quintana, S. M. (1998). Children's developmental understanding of ethnicity and race. *Applied and Preventive Psychology*, *7*, 27–45.

Quintana, S. M. (2008). Racial perspective taking ability: Developmental, theoretical, and empirical trends. In S. M. Quintana & C. McKown (Eds.), *Handbook of race, racism, and the developing child*. Hoboken, NJ: Wiley.

Reardon, S. F., Farrell, C. R., Matthews, S. A., O'Sullivand, D., Bischoff, K., & Firebaugh, G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research*, *38*, 55–70.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*, 853–891.

Reardon, S. F., & Robinson, J. P. (2007). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. F. Ladd (Ed.), *Handbook of research in education finance and policy*. New York: Routledge.

Rubovits, P. C., & Maehr, M. L. (1973). Pygmalion in Black and White. *Journal of Personality and Social Psychology*, *25*, 210–218.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *11*, 235–294.

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, *76*, 451–466.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, *59*, 7–13.

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, *16*, 125–141.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *227*, 918–924.

Seidman, E. (1983). Unexamined premises of social problem solving. In E. Seidman (Ed.), *Handbook of social intervention* (pp. 48–71). Beverly Hills, CA: Sage.

Slavin, R. E., & Madden, N. A. (2001). *Reducing the gap: Success for all and the achievement of African American students* Presented at the annual meeting of the American Educational Research Association. Seattle, WA.

Spencer, M. B. (1999). Social and cultural influences on school adjustment: The application of an identity-focused cultural ecological perspective. *Educational Psychologist*, *34*, 43–57.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 261–302). San Diego, CA: Academic Press.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Tharp, R. G. (1989). Psychocultural variables and constants: Effects on teaching and learning in school. *American Psychologist*, *44*, 349–359.

Thernstrom, A., & Thernstrom, S. (2002). *No excuses: Closing the racial gap in learning*. New York: Simon & Schuster.

Tyler, T. R., & Huo, Y. J. (2000). *How different ethnic groups react to legal authority*. San Francisco: Public Policy Institute of California.

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the national assessment of educational progress* (NCES No. 2009-455). Washington, DC: U.S. Department of Education.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, *20*, 1132–1139.

Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*. Cambridge, MA: Harvard University Press.

Weinstein, R. S., & McKown, C. (1998). Expectancy effects in context: Listening to the voices of students and teachers. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 7, pp. 215–242). Greenwich, CT: JAI Press.

Weinstein, R. S., Soulé, C. R., Collins, F., Cone, J., Mehlhorn, M., & Simontacchi, K. (1991). Expectations and high school change: Teacher-researcher collaboration to prevent school failure. *American Journal of Community Psychology*, *19*, 333–364.

Yeung, W. J., & Conley, D. (2008). Black-White achievement gap and family wealth. *Child Development*, *79*, 303–324.

Yeung, W. J., & Pfeiffer, K. M. (2009). The Black-White test score gap and early home environment. *Social Science Research*, *38*, 412–437.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1137–1144

Local Redundancy Governs Infants' Spontaneous Orienting to Visual-Temporal Sequences

Caspar Addyman and Denis Mareschal
University of London

Two experiments demonstrate that 5-month-olds are sensitive to local redundancy in visual-temporal sequences. In Experiment 1, 20 infants saw 2 separate sequences of looming colored shapes that possessed the same elements but contrasting transitional probabilities. One sequence was random whereas the other was based on bigrams. Without any prior exposure, infants looked longer at the random sequence. In Experiment 2, 17 infants looked equally long at bigram- and trigram-based sequences. However, an analysis of local redundancy revealed that in both experiments disengagement from the sequences was governed by local repetitions rather than by global sequence statistics. This finding suggests that a spontaneous sensitivity to stimulus complexity helps orient infants to sequences they can learn from.

It is well known that infants respond to stimulus complexity in static displays. In Cohen's (1972) seminal study on attention getting and attention holding, 4-month-olds' attention was held for longer by more complex visual stimuli (checkerboards with greater numbers of squares), supporting an earlier finding by Brennan, Ames, and Moore (1966). According to this view, a sensitivity to complexity might bootstrap infants' learning by helping them orient to particularly "interesting" stimuli. This is important because natural scenes possess substantial statistical redundancy and numerous correlations in space and time (Field, 1994), not all of which are informative. In spite of this, infants excel at learning from sequential stimuli. Artificial grammar learning (AGL) experiments show that infants can track transitional probabilities (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996) and learn other regularities in auditory stimuli (Gómez & Gerken, 1999; Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and in the visual domain (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Kirkham, Slemmer, Richardson, & Johnson, 2007). However, none of these studies have examined the complexity of the test materials as a performance factor in itself.

One reason for this gap in research is that complexity is a difficult concept to quantify. For the purposes of this article, we consider global complexity from an information theoretic point of view (Shannon & Weaver, 1949) and local redundancy as measured in terms of repetitiveness. Global complexity requires the infant to encode aspects of the sequence that are true across the sequence as a whole (e.g., the underlying grammar, the transitional probabilities, expected frequencies of co-occurrence), whereas local redundancy only requires the infant to encode information available within a short time window (e.g., the number of times a target item has appeared within the last four or five elements). Chater (1996) initially proposed that the ideas of simplicity and complexity could potentially provide a "unifying principle" for perception, learning, and other phenomena in cognitive science (Chater & Vitanyi, 2003). Hsu, Chater, and Vitanyi (2011) applied these ideas to child language acquisition, and a number of authors have applied similar ideas to infant word segmentation (Brent & Cartwright, 1996; Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Venkataraman, 2001). Most recently, Pothos (2010) applied these principles to adult AGL, demonstrating that the relative complexity of the training set and the test items could explain which ones are selected.

Complexity is especially relevant in information-rich domains like language learning, object recognition,

---

This research was supported by European Commission Framework 6 NEST Contract 516542 and UK ESRC-ORA RES-360- 25-056. We wish to thank Natasha Kirkham for useful discussion of this research, Sarah Cayless for her help coding the data, and the parents and children who took part in this research. Denis Mareschal is a current recipient of a Wolfson Royal-Society research merit award.

Correspondence concerning this article should be addressed to Caspar Addyman, Centre for Brain and Cognitive Development, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom. Electronic mail may be sent to c.addyman@bbk.ac.uk.

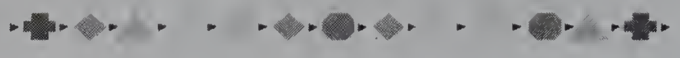
© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved.
0009-3920/2013/8404-0003
DOI: 10.1111/cdev.12060

or categorization in which it is difficult to determine what part of a stimulus is signal and what part noise. However, developmental research has largely ignored complexity as a predictive measure of stimulus preference or learning. An exception to this is the work of Son, Smith, and Goldstone (2008) showing that toddlers who learned with simple objects were able to generalize better according to shape than those who learned with complex objects. In a visual category learning task Robinson and Sloutsky (2007) found that increasing stimulus complexity by adding an auditory label increased the time infants spent looking at the stimulus, but reduced their category learning. In both these cases, increased complexity hindered learning. However, much research also shows that the rich and redundant cues in natural speech help rather than hinder infants to learn (e.g., Jusczyk & Aslin, 1995; Jusczyk & Thompson, 1978; Thiessen, Hill, & Saffran, 2005), thus posing somewhat of a conundrum.

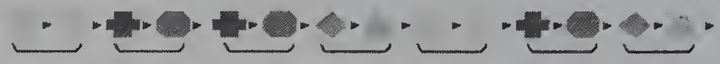
It is possible that perceptual and memory constraints could determine when complexity helps or hinders infant learning. Turkewitz and Kenny (1982) speculated that sensory limitations in early infancy might actually be adaptive, simplifying the perceived complexity of the world and facilitating learning. For example, Elman (1993) demonstrated theoretically that reduced working memory could potentially benefit a young language learner by allowing them to "start small," learning short distance case and clause agreements before long ones. Infants' performance may be highly dependent on the cognitive resources they have available to them. This suggests that it is important to consider the possible timescales over which statistical learning operates.

To explore these issues, we focused on Kirkham et al.'s (2002) seminal study of infant sequential learning. They habituated 2-, 5-, and 8-months-olds to a structured sequence of looming colored shapes. Six distinct shapes were assigned into three pairs (see Figure 1b). After habituation, the infants saw six test trials, three with the familiar pair-based structure and three with random, nonrepeating sequences (see Figure 1a). The trials alternated with the order counterbalanced and across the two test conditions the frequencies of the individual shapes were identical. Infants at all ages looked longer at the random sequence. Kirkham et al. took this as evidence of a domain general statistical learning mechanism operating from a very early age. A follow-up study also found a similar preference in newborns (Bulf, Johnson, & Valenza, 2011), although discrimination was only found in a "low-demand" condition with four distinct shapes.

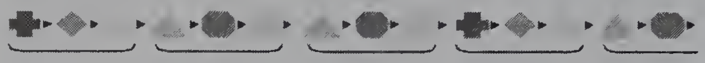
Stimuli



(a) Random Sequence, Each Item Is Followed by a Different Shape with Equal Probability.



(b) Pair-Based Grammar Where the Black Bracket Indicates the Fixed Deterministic Groupings.



(c) Triplet-Based Grammar

Figure 1. Sequence types used in this study. Experiment 1 compared infants' responses with (a) and (b). Experiment 2 compared infants' responses with (b) and (c). (a) Random sequence, each item is followed by a different shape with equal probability. (b) Pair-based grammar where the black bracket indicates the fixed deterministic groupings. (c) Triplet-based grammar.

Although those studies did demonstrate that even very young infants prefer a more random sequence after habituation to a statistically constrained pattern, neither experiment examined whether infants have a prior preference for a less predictable sequence. There are several reasons to suspect that they would; the structured sequence is more likely to contain highly repetitive runs of alternating items and contains far fewer pairwise transitions than the random sequence. Infants could be responding to local redundancy than the global sequence statistics. Thus, this study aims to investigate infants' prior preferences and response to local sequence redundancy.

Experiment 1 follows as closely as possible Kirkham et al.'s (2002) methodology, but omits the habituation phase. We chose 5-month-old infants because these were the middle group in the original study. However, unlike Kirkham et al., we allowed each test trial to last up to 90 s to allow a greater opportunity for discrimination.

Experiment 1: Random Versus Pair-Based Sequences

Method

Participants. Twenty infants (10 females) with a mean age of 154 days (range = 144–162) were tested. The participants were full-term infants who had experienced no birth complications and with no history of familial color blindness. A further 11 infants were tested but were excluded due to fussiness

(5), equipment failure (3), experimenter error (2), or parental interference (1). Infants were from a mix of ethnicities and backgrounds, but the sample was predominantly European and middle class. Infants were recruited from a pool of volunteer families who were part of the center's participant database. Parents were not paid, but travel expenses were reimbursed and infants were given a certificate and a small gift.

Apparatus. A Macintosh G4 computer running MATLAB R2006a (Mathworks Inc., Natick, MA) was used to control the experiment and stimuli were displayed on a 49-cm Panasonic color monitor. Video of the infant was recorded using a low-light DV camera positioned below the screen, facing the infant. These were situated in a soundproofed testing booth.

Stimuli. The stimuli consisted of six colored shapes (turquoise square, blue cross, yellow circle, pink diamond, green triangle, and red octagon) presented in sequence on a black background. As in Kirkham et al. (2002) shapes loomed on the screen, increasing from 4 to 24 cm in height (2.4°–14.6°) over the course of a second. Shapes were presented one after the other with no interval between them.

In the random condition, the shapes were presented in a different random sequence each time, constrained so that each shape appeared with equal probability, but no shape appeared twice in a row. In the pairs condition, stimuli were assigned to three fixed pairs which repeated randomly (see Figure 1). Each infant saw a total of six trials, three structured and three random. For any given infant, presentation alternated between structured and random sequences and order was counterbalanced so that half the infants saw the structured sequence first, the other half saw the random sequence first.

Procedure. Infants were seated in a car seat, in a darkened room, at a distance of 95 cm from the screen. The caregiver remained in the room but was seated behind and out of view of the infant and was instructed not to interact with the infant.

An observer viewed the infant on a separate monitor in the main lab, and started and finished each trial on the basis of the individual infants' looking behavior. The observer depressed the space bar on the computer whenever the infant was attending, releasing it if the infant looked away. Each trial finished if the computer determined that the infant had looked away continuously for 2 s or if the trial had lasted 90 s. Trials always finished when the current shape had finished looming. Between trials the screen showed a small blinking fixation square and, if necessary, infants were reoriented to the screen by an auditory attention getter.

Results

Offline manual coding of the videos produced two data sets for analysis. The first was a looking time measure per trial for each infant, calculated by adding up all the time that the infant was looking at the screen per trial. A second experimenter, blind to the hypothesis, double-coded a randomly selected 25% of the trials. The Pearson correlation between the first and second coders' judgments of looking time was r = .981, p = .001. The second data set composed of the exact sequence of shapes seen by the infant on a given trial, allowing for any short looks away. Infants were deemed not to have seen a particular stimulus shape if they saw less than either the first 300 ms or last 100 ms of that stimulus. This latter data set allowed us to explore infants' response as a function of local redundancy in the sequences.

Looking time analyses. For the purposes of analysis, the six trials were divided into three blocks each consisting of a pair of sequences: one structured, one random (Figure 2). To investigate whether the type of sequence had an influence on the length of looking, a two-within, one-between mixed analysis of variance (ANOVA; 3 × 2 × 2, Block × Pattern × Order) was performed. The ANOVA revealed a highly significant main effect of block, $F(2, 36) = 14.8$, $p = .001$, generalized $\eta^2 = .23$. Infants' looking times decreased across the three pairs of sequence presentations, indicating that they became familiar with the overall experimental setup. In line with our hypothesis, there was a significant main effect of pattern, $F(1, 18) = 4.72$, $p = .043$, generalized $\eta^2 = .017$. Infants looked longer on

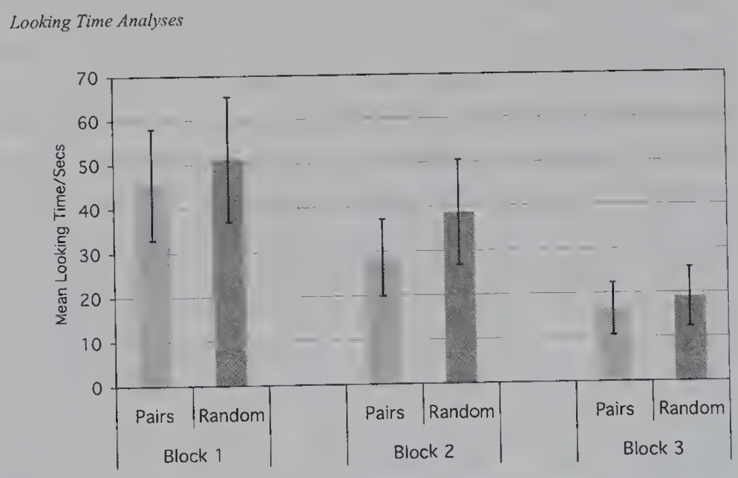


Figure 2. Mean looking times in seconds toward random and pair-based sequences by 5-month-olds. Error bars are 95% confidence intervals, $N = 20$.

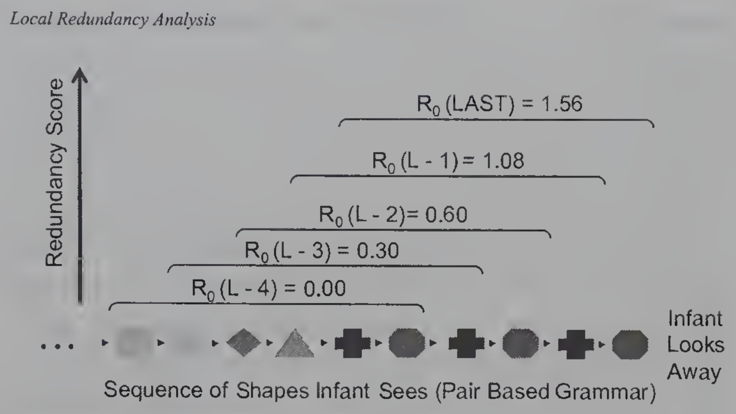


Figure 3. Local redundancy calculated over a moving six-item window. Scores are higher when shapes repeat.

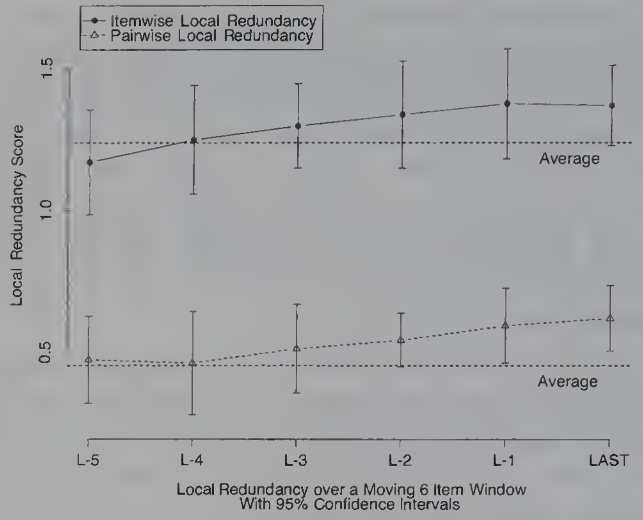


Figure 4. Local redundancy analysis for Experiment 1. The graph shows the mean redundancy scores for each of the last six items presented (L-5 to LAST) prior to the infant looking away from the sequences, calculated with a six-item window. The upper line shows the itemwise local redundancy and the lower line the pairwise redundancy. The dotted lines show the average local redundancy across all items presented. $N = 20$.

average at the random pattern in each presentation trial. There was no interaction with block, indicating that the preference is consistent across the three blocks and was present from the first pair of trials. Moreover, order of presentation was not a significant factor.

Local redundancy analysis. An alternative analytic approach is to identify when infants choose to look away in a sequence; that is, do they look away when there is high local redundancy? For example, infants may look away when they see a run of items that goes . . .-blue-red-blue-red-blue-red (Figure 3). Repetitive intervals are more likely in constrained grammars and therefore on average infants may look less at these due to the transient presence of more local redundancy in such sequences. Consequently, we analyzed the results of Experiment 1 in terms of local redundancy, adapting a measure introduced by Jamieson and Mewhort (2005). (See online supporting information Appendix S1.)

Mean item and pairwise local redundancy scores were calculated looking back over the previous 6 items from the last item the infant looked at, the last but one item, and so on, averaged across all trials (Figure 4). We used a 6-item window as there were six different shapes and much smaller or larger windows would not capture a local effect. Moreover, trials averaged 33 items in length so 6-item window provides a good contrast with global effects. Analyses based on 8- and 4-item windows showed a similar pattern of results. The graphs slope upward, suggesting that infants looked away when local redundancy increased. This was confirmed with repeated measures ANOVA with distance from end of trial as the repeated measure (six levels). There was a significant effect for itemwise local redundancy $F(1, 118) = 4.15$, $p = .044$, generalized = .035, and for pairwise local redundancy, $F(1, 118) = 4.25$, $p = .042$, generalized = .036. These results show: (a) that infants are sensitive to both item and pairwise redundancy, and (b) that they are more likely to look away when the sequences become more locally redundant.

Discussion

Five-month-olds spontaneously looked longer at a random temporal sequence compared to a sequence of bigrams. Frequencies of individual items were equal, so infants could only discriminate these sequences on the basis of pairwise transitions or some measure of sequence complexity. They discriminated spontaneously from the first block without any prior exposure to either sequence type. This suggests that a process simpler than global statistical learning is in operation. Results from the redundancy analysis indicate that infants' performance is determined by short-term repetitions in the stochastic sequences presented to the individual infants.

However, infants could still have very rapidly learned some aspects of the grammar, perhaps a single pairing of shapes, the repetition of which leads them to disengage from that particular sequence. These two possibilities are confounded in the current experiment because the pairwise grammar has more repetition than the random grammar. If the infants respond more to local redundancy than to grammar, then we would predict that in a new task where both sequences have a grammatical structure, there would be no difference in looking time between conditions but local redundancy scores would still predict when infants looked away. Experiment 2 aims to resolve this issue.

Experiment 2: Pairs Versus Triplets

In Experiment 2 we compare the pairs sequence from Experiment 1 with a more predictable sequences consisting of the same shapes arranged in two triplets. If infants look longer at globally complex sequences, then, in this case, they should look longer at the pair-based sequence. If, however, they are responding to local repetition, then disengagement should be predicted by an increase in the redundancy scores at the end of the trials.

Method

Participants. A total of 17 infants (8 females) completed the experiment; their mean age was 154 days (range = 137–163). A further 2 were excluded due to experimental error and equipment failure, and 8 were excluded due to fussiness. Infants were screened and recruited as in Experiment 1.

Procedure and stimuli. The procedure was identical to Experiment 1. The six trials alternated between three sequences consisting of three pairs, and three sequences consisting of two triplets of shapes. Figures 1(b) and (c) illustrate the types of sequence used in this experiment.

Results

The videos were coded as before and a second experimenter blind to the hypothesis double-coded a randomly selected 25% of the data. The Pearson correlation between the coders' scores was $r = .989$, $p = .001$. Looking times are plotted in Figure 5.

Looking time analysis. We conducted the same analysis as in Experiment 1. There was a highly significant main effect of block, $F(2, 30) = 21.5$, $p = .001$, generalized = .29. This was due to the general decrease in looking throughout the experiment. There was also a significant Pattern × Order interaction, $F(1, 15) = 11.1$, $p = .005$, generalized = .066. This interaction is a consequence of the general decline in looking over the trials irrespective of pattern; half the infants saw triplets followed by pairs, T1-P1-T2-P2-T3-P3, whereas the others saw the opposite order, P1-T1-P2-T2-P3-T3. Therefore, due to the global decline in looking across trials, Tx > Px for the former infants whereas Px > Tx for the latter. Importantly, there was no main effect of pattern, $F(1, 15) \ll 1$. There were no other significant main effects or interactions. Infants do not look longer at either the sequences generated by a pair-based or a triplet-based grammar.

Local redundancy analysis. An identical set of complexity analyses were carried out as in Experiment 1. The results showed the same pattern as before (Figure 6); infants appear to look away at points when the sequence is more repetitive and locally redundant than average. As before, we confirmed this with a repeated measure ANOVA with distance from end of trial as the repeated measure (six levels). There were significant effects for both item-wise local redundancy $F(1, 100) = 9.25$, $p = .003$, generalized = .092, and for pairwise local redundancy $F(1, 100) = 13.5$, $p = .001$, generalized = .048. Thus, infants were sensitive to both item- and pairwise local redundancy. They were more likely to look away when the sequences became more locally repetitive.

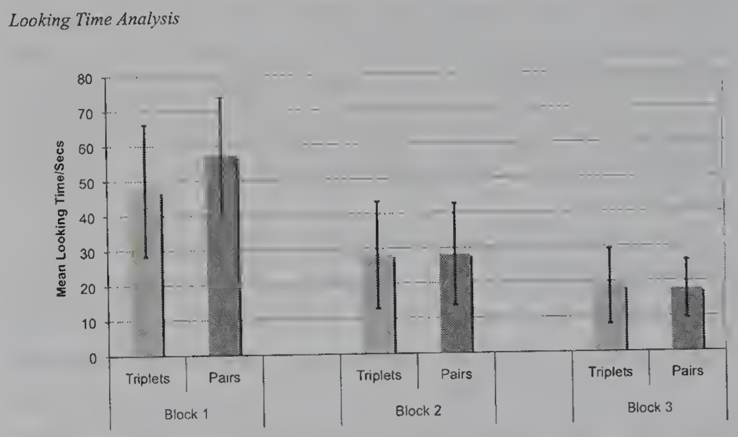


Figure 5. Mean looking times in seconds toward pair- and triplet-based sequences by 5-month-olds. $N = 17$.

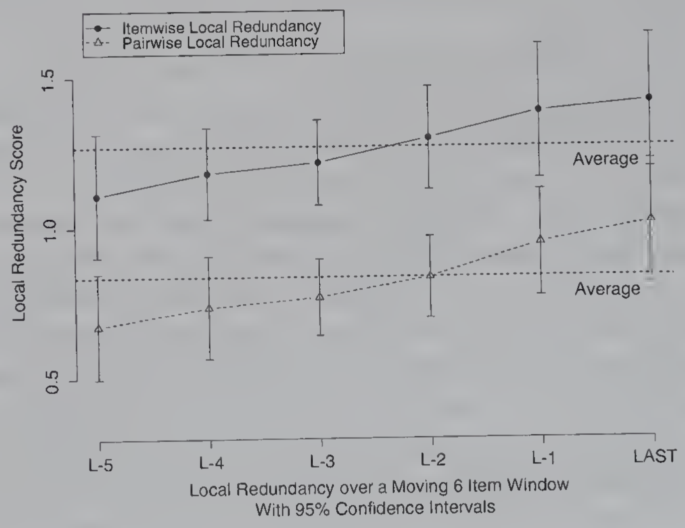


Figure 6. Local redundancy scores for Experiment 2. $N = 17$.

Discussion

In Experiment 2 infants looked equally long at pair- and triplet-based grammars. This does not support the hypothesis that infants would look longer at the globally more complex (pair-based) sequence. However, as in Experiment 1, it was found that single-item and pairwise redundancy measures were significantly above average at the point when infants looked away. Thus, infants were more likely to look away when the most recent few items they had seen were more repetitive.

General Discussion

We conducted two experiments exploring 5-month-olds' responses to visual-temporal sequences with no prior learning or familiarization. Taken together they provide strong evidence that young infants are sensitive to complexity and will spontaneously prefer more complex, less orderly grammatical sequences. A detailed local redundancy analysis of the items and pairs the infants saw at the point they disengaged from these sequences supported the idea that the infants were not learning global properties of the grammars presented, but were responding to local redundancy over a six-item window.

First, these finding suggest a possible alternative interpretation to the results of Kirkham et al. (2002) and Bulf et al. (2011). This study shows that there need not have been any statistical learning in those studies. Infants may, instead, have spontaneously detected differences in local redundancy between the grammatical and random sequences. Performance in all cases could be explained as response to relatively localized simple features rather than sophisticated global statistical or grammar learning.

It is worth emphasizing that these findings do not in themselves undermine other results showing statistical learning in infants. A great deal of research shows that infants can use frequency information and transition probabilities to segment sequential stimuli (Fiser & Aslin, 2002; Kirkham et al., 2007; for a review and new model of sequence learning, see French, Addyman, & Mareschal, 2011). However, this study underscores the need for direct measures of *learning* in these types of studies as discrimination can potentially be attributed to differences in local redundancy (see also Pothos, 2007). For example, some studies have used counterbalancing and single-item test cases to control for alternative explanations (e.g., Aslin et al., 1998; Pelucchi, Hay, & Saffran, 2009a,2009b). In fact, Experiment 4 of Kirkham et al. (2007) demonstrated reduced eye movement latencies to statistically more likely spatial transitions. There is even direct evidence of statistical learning in newborns using ERP measures (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Thus, the novel contribution of this study is less about striking a cautionary note when exploring statistical learning, but more to reiterate the importance of the question raised by Cohen (1972) and more recently Chater and Vitanyi (2003). Namely, how does a spontaneous sensitivity to stimulus complexity interact with the basic attentional mechanisms that direct perception or learning?

Cohen (1972) found that increasing complexity of a static visual stimulus predicted attention holding. Our results clearly demonstrate that considerations of local redundancy are strongly predictive of infants' disengagement from a display. This merits further investigation in a wider range of contexts. In particular, it suggests that more links should be made between the study of infant learning as a statistical associative process and the cognitive constraints governing that process. Stimulus complexity has been largely overlooked in previous infant AGL research.

Our results are most straightforwardly explained in terms of infant sensitivity to local redundancy (Jamieson & Mewhort, 2005) in the stimulus stream. This supports the findings of Haith and colleagues (Adler & Haith, 2003; Haith, 1993; Haith, Adler, & Was, 1996) that infants as young as 3 months will learn to anticipate regularities in spatiotemporal sequences over a window of a few seconds. Similarly, just as Pothos (2010) demonstrates a role for complexity in adult AGL research, the present results suggest an analogous importance in the interpretation of infant AGL results and potentially more widely. This is one of the first studies, along with Kidd, Piantadosi, and Aslin (2012), to demonstrate that complexity or simplicity principles that operate in adult cognition also hold in early infancy.

In conclusion, this study shows that infants have a spontaneous sensitivity to complexity of visual-temporal sequences. Infants' disengagement from the stimulus was best explained in terms of an awareness of local redundancy and not global (grammatical) structure. This finding offers a more parsimonious explanation of some infant statistical learning results, reopens an old debate in the infancy literature on stimulus complexity, and underscores the centrality of complexity in early cognition.

References

Adler, S., & Haith, M. (2003). The nature of infants' visual expectations for event content. *Infancy*, *4*, 389–421.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month old infants. *Psychological Science*, *9*, 321–324.

Brennan, W. M., Ames, E. W., & Moore, R. W. (1966). Age differences in infants' attention to patterns of different complexities. *Science*, *151*, 354–356.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*, 127–132.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*, 566–581.

Chater, N., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, *43*, 869–879.

Elman, J. (1993). Learning and development in neural networks—The importance of starting small. *Cognition*, *48*, 71–99.

Field, D. (1994). What is the goal of sensory coding? *Neural Computation*, *6*, 559–601.

Fiser, F., & Aslin, R. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822–15826.

Frank, M., Goldwater, S., Griffiths, T., & Tenenbaum, J. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*, 107–125.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*, 614–636.

Gómez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.

Haith, M. (1993). Future-oriented processes in infancy, the case of visual expectations. In C. E. Granrud (Ed.), *Visual perception and cognition in infancy*, (pp. 235–264). Hillsdale, NJ, Erlbaum.

Haith, M., Adler, S., & Was, T. (1996). Time factors in infant expectation. *Infant Behavior and Development*, *19*, 157

Hsu, A. S., Chater, N., & Vitanyi, P. M. B. (2011). The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition*, *120*, 380–390.

Jamieson, R., & Mewhort, D. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 9–23.

Jusczyk, P., & Aslin, R. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Jusczyk, P., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Attention, Perception, & Psychophysics*, *23*, 105–109.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*, e36399.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*, 1559–1571.

Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.

Pelucchi, B., Hay, J., & Saffran, J. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*, 244–247.

Pelucchi, B., Hay, J., & Saffran, J. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*, 674–685.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*, 227–244.

Pothos, E. (2010). An entropy model for artificial grammar learning. *Frontiers in Psychology*, *1*, 1–13.

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, *11*, 233–253.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Son, J., Smith, L., & Goldstone, R. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*, 626–638.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21–29.

Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*, 53–71.

Turkewitz, G., & Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, *15*, 357–368.

Venkataraman, A. (2001). A statistical model for word discovery in transcribed speech. *Computational Linguistics*, *27*, 351–372.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Measuring Local Redundancy.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1145–1151

Longitudinal Relations of Intrusive Parenting and Effortful Control to Ego-Resiliency During Early Childhood

Zoe E. Taylor, Nancy Eisenberg, and Tracy L. Spinrad
Arizona State University

Keith F. Widaman
University of California, Davis

Longitudinal relations among ego-resiliency (ER), effortful control (EC), and observed intrusive parenting were examined at 18, 30, and 42 months of age (*N*s = 256, 230, and 210) using structural equation modeling. Intrusive parenting at 18 and 30 months negatively predicted EC a year later, over and above earlier levels. EC at 30 months mediated the negative relation between 18-month intrusive parenting and ER at 42 months when controlling for stability of the variables. ER did not predict EC. The findings suggest that intrusive parenting may have a negative effect on children's ego-resiliency through its effects on children's abilities to regulate attention and behavior.

A key component of healthy adjustment is psychological flexibility, or the dynamic process by which individuals adapt to fluctuating situational demands, reconfigure mental resources, shift perspectives, and balance competing desires and needs (Kashdan & Rottenberg, 2010). A personality characteristic associated with psychological flexibility is ego-resiliency (ER). ER reflects adaptability to environmental stress and change (Block & Block, 1980). Individuals with high ER are able to adapt to changing circumstances, shift behaviors as needed, and use problem-solving strategies flexibly. A nonresilient individual displays little adaptive flexibility, is disquieted by change, has the tendency to perseverate or become disorganized when dealing with stress, and has difficulty recouping after traumatic experiences (Block & Block, 2006). Supporting this notion, ER has been associated with diverse positive developmental outcomes such as social competence (Block & Block, 1980), low externalizing and internalizing symptoms (Chuang, Lamb, & Hwang, 2006; Hofer, Eisenberg, & Reiser, 2010), and cognitive functioning and attentiveness (Martel et al., 2007).

Consequently, examining how ER develops and is fostered in early childhood has important implications for later adjustment. Personality characteristics reflect cognitive structures and adaptive strategies that develop across time through socialization and further expression of temperamental characteristics (Rothbart & Bates, 2006). Therefore, it is likely that ER is shaped by early developing temperamental traits as well as by environmental influences such as parenting. However, relatively little is known about how ER develops from an early age or how it is linked across time with other developmental processes.

Effortful control (EC) is a temperamental characteristic that likely contributes to the early development of ER. EC includes the ability to voluntarily or willfully focus and shift attention and inhibit or initiate behaviors, to plan, and to detect errors (Rothbart & Bates, 2006). These abilities can be viewed as tools that help modulate emotion and behavior; consequently, individuals with high EC would be expected to have an advantage in regard to adapting effectively in stressful situations (Eisenberg et al., 2004). Although EC and ER have been positively linked in samples of preschoolers (Cumberland-Li, Eisenberg, & Reiser, 2004), elementary school children

---

Zoe E. Taylor and Nancy Eisenberg, Department of Psychology, Arizona State University; Tracy L. Spinrad, School of Social and Family Dynamics, Arizona State University; Keith F. Widaman, Department of Psychology, University of California, Davis.

This project was funded by National Institute of Mental Health Grant 5 R01 MH60838 to Nancy Eisenberg and Tracy L. Spinrad. We would like to thank the children, parents, and teachers involved in this study. In addition, we appreciate our undergraduate research assistants for their contributions.

Correspondence concerning this article should be addressed to Zoe E. Taylor, Department of Psychology, Arizona State University, Tempe, AZ 85287-3701. Electronic mail may be sent to zetaylor@asu.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0004
DOI: 10.1111/cdev.12054

(Eisenberg et al., 2003; Eisenberg et al., 2004; Eisenberg et al., 2010), and adolescents (Hofer et al., 2010; Martel et al., 2007), these relations have not been examined in toddlerhood.

The social environment also is likely to affect the development of children's personality and regulatory traits. Exposure to intrusive parenting during early childhood may be particularly detrimental because this is a period when many important developmental changes, such as increased regulation and individual differences in personality, are emerging (Edwards & Liu, 2002; Sturge-Apple, Davies, Martin, Cicchetti, & Hentges, 2012). Intrusive, insensitive, or controlling parents likely undermine the development of independent coping skills and discourage children's independent behaviors, whereas sensitive parents may model effective coping skills, provide appropriate support when children are overwhelmed, and expose children to manageable, age-appropriate stressors (Power, 2004). Furthermore, when mothers are unresponsive, punitive, or insensitive, children may experience heightened arousal that disrupts their ability to self-regulate, whereas mothers who are supportive and sensitive likely model constructive ways to manage stress and relationships (Eisenberg et al., 2010).

Although researchers have not specifically examined the relations between intrusive parenting and ER, lower levels of ER have been found in toddlers who display contradictory attachment behaviors when under stress (Smeekens, Riksen-Walraven, & van Bakel, 2009) and in maltreated school-age children (Shonk & Cicchetti, 2001). Conversely, supportive parenting has been found to foster children's ER (Block & Block, 1980; Stams, Juffer, van IJzendoorn, & Hocksbergen, 2001). Similarly, insensitive, punitive parenting is inversely related to children's EC whereas reversed findings have been found for warm parenting (Eisenberg et al., 2010; Graziano, Keane, & Calkins, 2010). However, punitive versus warm parenting is not identical to intrusive parenting—for example, parents can be warm and intrusive—so one cannot assume that intrusive parenting undermines ER (Denissen, van Aken, & Dubas, 2009).

The present study investigated the longitudinal relations among intrusive parenting, ER, and EC. We expected negative relations between intrusive parenting and ER, and hypothesized that EC would mediate those relations across time. We expected that EC, because it is a fairly rudimentary temperamental capacity, to predict ER, which is viewed as an aspect of personality and expected to develop from temperament and experience dealing with stress. ER has seldom been examined in young children, especially in relation to EC and intrusive parenting. Furthermore, unlike in prior studies involving one or two assessments (e.g., Eisenberg et al., 2004), we extended prior research by examining the relations between ER and EC across three assessments, which is desirable for testing mediated relations (Maxwell & Cole, 2007), the stability of constructs, and bidirectional paths.

Method

Sample and Procedure

Participants were families in an urban area, assessed at approximately 18 (T1), 30 (T2), and 42 (T3) months of age ($ns = 256$, 230, and 210). Children at T1 (55% male; M age = 17.78, $SD = 0.52$) were mostly Caucasian and not Hispanic (87%) or Hispanic (13%). Mean parental education was some college and mean annual family income was \$45,000–\$65,000. Mothers who attrited were younger at the child's birth ($M = 27.16$ years) than were mothers still participating at T3 ($M = 29.70$), $t(243) = 2.93$, $p < .01$. No other key variables were related to attrition. At each assessment, mothers and nonparental caregivers (e.g., babysitter, preschool teacher; $ns = 176$, 153, 151) were mailed questionnaires to mail back or to bring to the laboratory. Mothers and children participated in laboratory sessions lasting 1.5–2 hr.

Measures

Children's ER. Mothers and nonparental caregivers rated children's ER (1 = *highly undescriptive*, 9 = *highly descriptive*) at all time points using an 11-item adapted questionnaire version of Block and Block's (1980) Q-sort. This shortened scale was constructed by Eisenberg and colleagues (Eisenberg et al., 1996; Eisenberg et al., 2003) to reflect a purer version of resiliency that did not overlap with other constructs, such as negative emotionality. One item was dropped ("Can talk about unpleasant things that have happened to him/her") as it was not asked at 18 and 30 months as it was inappropriate for the ages assessed. Items included "Can bounce back or recover after a stressful or bad experience." In order to measure ER at a latent level, items from both reporters were randomly composited into three parcels as recommended by Kishton and Widaman (1994). Alphas for combined reporters = .67,

.75, and .77. Indicators were significantly correlated at all time points.

Children's EC. Mothers and nonparental caregivers rated EC (1 = *never,* 7 = *always*) using three 12-item subscales (attention-focusing, attention-shifting, and inhibitory-control) of the Early Childhood Behavioral Questionnaire (Putnam, Gartstein, & Rothbart, 2006) at T1 and T2, and the Child Behavior Questionnaire (Rothbart, Ahadi, Hershey, & Fisher, 2001) at T3, for example, "When engaged in play with his/her favorite toy, how often did your childstay involved for 10 minutes or more?" Reporters were combined for each subscale (inhibitory-control αs = .86, .88, and .86; attention-focus αs = .79, .82, and .77; and attention-shifting αs = .75, .74, and .74). Indicators were significantly correlated at all time points.

Observed intrusive parenting. Coders assessed mothers' behavior at T1, T2, and T3 from videotapes of mother–child interactions during a teaching task, a free-play task, and a clean-up task, each lasting 3 min. In the teaching task, mothers taught their child to complete a puzzle (T1, T2) or Lego model (T3) with whatever strategies they would use at home. During the free-play task, mothers were given a basket of toys and asked to play as they would at home. Mothers' behavior was rated for intrusive/overcontrolling behavior every 30 s (1 = *no intrusive behavior,* 4 = *extreme intrusive behaviors*). Intraclass correlation coefficients (ICCs), assessed for 23%–30% of the sample, were .81, .71, and .83 for teaching and .81, .86, and .83 for free play. During the clean-up task, mothers were asked to have their children pick up the toys as if they were at home. Controlling behavior was rated every 15 s (0 = *absent,* 1 = *present*; ICCs for 26% of the sample = .70, .82, and .85). Examples of intrusive behavior included offering a continuous barrage of toys or not allowing children to touch an object they were reaching for.

Control variables. Control variables included maternal education (1 = *grade school* to 7 = *PhD or MD*); mothers' marital status (0 = *two-parent family,* 1 = *single-parent family*), child sex (0 = *male,* 1 = *female*), and household income (1 = < *\$15,000* to 7 = *over \$100,000*). Controls were analyzed on all latent variables and retained if they were significant.

Results

Preliminary Analyses

We examined zero-order correlations (Table 1) and then used structural equation modeling with full information maximum likelihood estimation (Mplus, Version 6; Muthén & Muthén, 2007) to evaluate the model (Figure 1). Because two indicators exhibited non-normality in skew or kurtosis–intrusiveness during the puzzle task at 18 and 30 months, we used MLR estimation, a modified form of maximum likelihood estimation that is robust to non-normality. First, we fit a confirmatory factor model with latent and control variables freely correlated. Unique variances of the variables were

Table 1
Zero-Order Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Resiliency T1	1.00												
2. Ef. Cont. T1	.30**	1.00											
3. Intr. Par. T1	−.18**	−.08	1.00										
4. Resiliency T2	.35**	.09	−.03	1.00									
5. Ef. Cont. T2	.19**	.57**	−.22**	.26**	1.00								
6. Intr. Par. T2	−.04	−.08	.57**	−.10	−.20**	1.00							
7. Resiliency T3	.31**	.13	−.12	.44**	.21**	−.15*	1.00						
8. Ef. Cont. T3	.15*	.39**	−.30**	.16*	.60**	−.33**	.31**	1.00					
9. Intr. Par. T3	−.08	−.14	.47**	−.13	−.19*	.55**	−.14	−.30**	1.00				
10. Mo. Ed. T1	.09	.08	−.36**	.16*	.08	−.29**	.09	.23**	−.36**	1.00			
11. Mar. status T1	.06	−.00	−.33**	.20**	.11	−.34**	.01	.10	−.39**	.32**	1.00		
12. Child Sex	−.01	−.08	.01	−.01	−.06	.20**	−.06	−.13	.07	.02	.01	1.00	
13. Income T1	.05	.05	−.23**	.12	.06	−.15*	−.08	.14	−.24	.50**	.38**	.12	1.00
M	6.76	4.16	0.99	6.88	4.46	0.87	6.89	4.42	1.01	4.21	0.84	0.56	4.08
SD	0.72	0.51	0.24	0.85	0.54	0.18	0.89	0.50	0.19	1.05	0.36	0.50	1.74

Note. Ef. Cont. = effortful control; Intr. Par. = intrusive parenting; Mo. Ed. = mother's education; Mar. status = marital status; T1 = 18 months; T2 = 30 months; T3 = 42 months.
$^{*}p < .05$. $^{**}p < .01$.

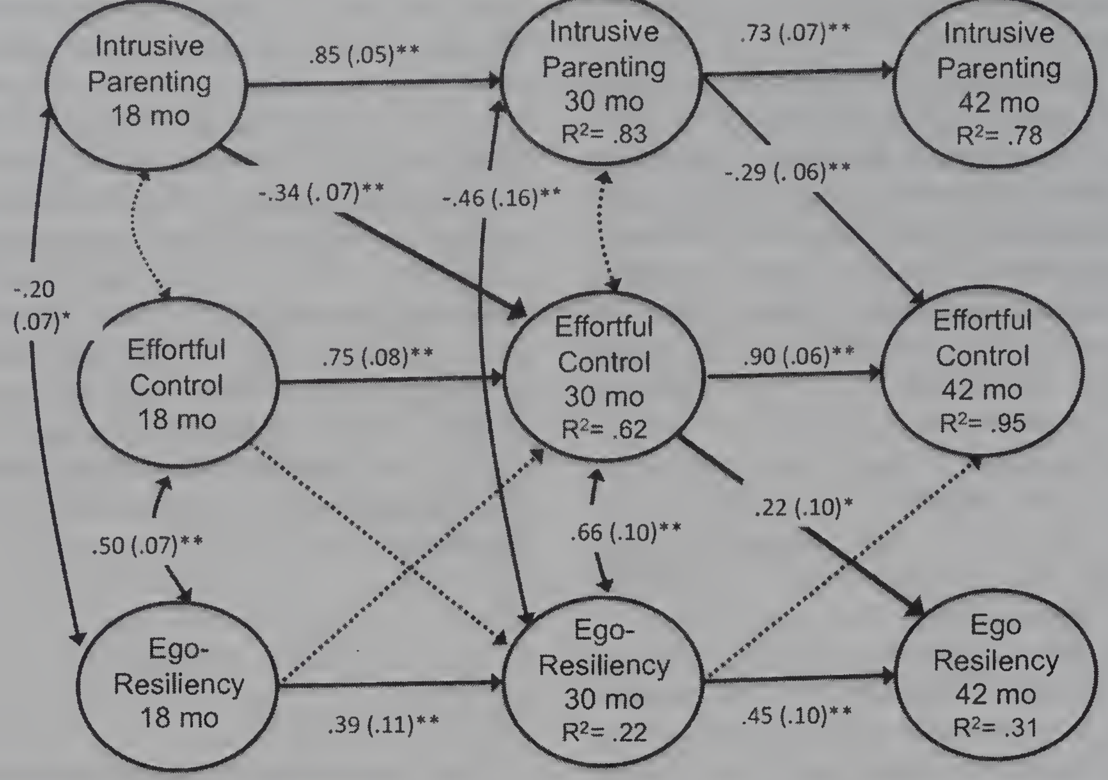

Intrusive Parenting 18 mo
.85 (.05)**
Intrusive Parenting 30 mo R²= .83
.73 (.07)**
Intrusive Parenting 42 mo R²= .78
-.34 (.07)**
-.46 (.16)**
-.29 (.06)**
-.20 (.07)*
Effortful Control 18 mo
.75 (.08)**
Effortful Control 30 mo R²= .62
.90 (.06)**
Effortful Control 42 mo R²= .95
.50 (.07)**
.66 (.10)**
.22 (.10)*
Ego-Resiliency 18 mo
.39 (.11)**
Ego-Resiliency 30 mo R²= .22
.45 (.10)**
Ego Resiliency 42 mo R²= .31

Figure 1. Results of the structural equation model. $\chi^2(402, N = 256) = 564.98, p = .99$; RMSEA = .04, CFI = 0.91, TLI = 0.90. Results are standardized coefficients with standard errors shown in parentheses. Dotted lines are nonsignificant. 18 mo = T1; 30 mo = T2; 42 mo = T3. Significant control variables: Education on ego-resiliency T1 and T2 and parenting T1 and T3; Marital status on parenting T1, and ego-resiliency T2 and T3; Child sex on parenting T2.
$^*p < .05$ (two-tailed test). $^{**}p < .01$.

allowed to covary within a construct when indicated by the modification indices. This measurement model demonstrated acceptable fit, $\chi^2(340, N = 256) = 497.53, p < .01$; root mean square error of approximation (RMSEA) = .04, comparative fit index (CFI) = .91, Tucker–Lewis Index (TLI) = .88. We next tested the invariance of the loadings for indicators across time to examine whether the relations of the latent variables to the manifest variables were constant. Given that we we used MLR estimation, we used the Satorra-Bentler (S-B) scaled (mean-adjusted) chi-squares, wherein the usual normal-theory chi-square statistic is divided by a scaling correction to better approximate chi-square under non-normality. Two constraints resulted in a deterioration of fit: the resilience B indicator, S-B $\chi^2(2, N = 256) = 10.58, p < .01$, and control from intrusive parenting, S-B $\chi^2(2, N = 256) = 9.04, p < .01$. However, the fit indices remained the same. Given the importance of maintaining continuity across the constructs, we retained all the loading constraints.

Structural Equation Analysis

We then specified a restricted structural model (Figure 1) and tested a series of paths. The paths from parenting to ER, as well as from EC to parenting, were nonsignificant and removed to reduce the complexity of the model. We left the path from ER to EC in the model in order to examine the direction of effects between EC and ER. To determine if there were differences in prediction from our constructs at different ages, we then placed a series of constraints on our model to test invariance among the paths in our model. Constraining stability path coefficients for each construct (e.g., T1 ER to T2 ER and T2 ER to T3 ER) as well as the concurrent correlation pathways between constructs (e.g., between EC and ER at T1 and again at T2), did not result in a significant change of fit, S-B $\chi^2(6, N = 256) = 3.98, p = .68$; thus, these constraints were retained. We then constrained cross-lagged paths. Lagged paths from parenting to EC and from ER to EC also did not result in a significant change of fit, S-B $\chi^2(2, N = 256) = 1.55, p = .46$, and these constraints were retained. However, lagged paths from EC to ER resulted in a significant deterioration of fit, S-B $\chi^2(1, N = 256) = 6.56, p < .01$, and these paths were left unconstrained. This final model demonstrated an adequate fit: $\chi^2(402, N = 256) = 564.98, p = .99$; RMSEA = .04, CFI = 0.91, TLI = 0.90. Factor loadings for the latent variables were all significant at $p < .01$ (range = .40 to .88; Table 2).

Table 2
Factor Loadings for Latent Variables, Standardized

Latent variables	T1 (18 mo) Est. (*SE*)	T2 (30 mo) Est. (*SE*)	T3 (42 mo) Est. (*SE*)
Ego-resiliency			
Resiliency A	.68 (.04)	.74 (.04)	.71 (.04)
Resiliency B	.72 (.06)	.79 (.03)	.89 (.04)
Resiliency C	.55 (.04)	.64 (.04)	.68 (.04)
Effortful control			
Attention-focus	.55 (.06)	.59 (.05)	.59 (.06)
Attention-shifting	.61 (.05)	.63 (.05)	.58 (.04)
Inhibitory-control	.57 (.05)	.59 (.05)	.70 (.07)
Intrusive parenting			
Intrusiveness, puzzle task	.45 (.05)	.48 (.05)	.40 (.04)
Intrusiveness, freeplay task	.58 (.05)	.60 (.07)	.44 (.07)
Control, clean-up task	.58 (.07)	.54 (.06)	.42 (.05)

Note. All factor loadings are significant at $p < .01$. Est. = estimate.

Intrusive parenting, EC, and ER were stable across all assessments (particularly EC and parenting). ER and EC were positively correlated within T1 and within T2, as were ER and parenting. EC and parenting were not correlated at any concurrent time point in the model, although they were correlated in the zero-order correlations at T2 and T3. Across time, intrusive parenting predicted EC at both time points. T1 EC did not predict T2 ER; however, there was a significant path from T2 EC to T3 ER. Both across-time paths from ER to EC were nonsignificant. Lastly, T2 EC mediated the relation between T1 intrusive parenting and T3 ER ($b = -.08$, $SE = .04$, $z = -2.00$, $p < .05$) using the model indirect test in Mplus.

Discussion

We tested a three-wave model with autoregressive and bidirectional cross-lagged paths to assess relations among young children's ER, EC, and intrusive parenting. Despite significant stability in all three constructs and correlations among some of the constructs within T1 and within T2, we found that EC mediated the relation of intrusive parenting to ER. Mediation suggests the effects of the intrusive parenting on ER were due to its effects on the development of effort control, which in turn provided skills used to become more ego-resilient over time. Because the panel model included controlling for prior levels of the variables, the paths can be interpreted as predicting change over time. Consistent with research with older children (e.g., Cumberland-Li et al., 2004; Eisenberg et al., 2010; Martel et al., 2009), we found concurrent associations between ER and EC. In addition, we found evidence that 30-month EC positively predicted 42-month ER, controlling for prior levels of ER and concurrent correlations between the two constructs. Given the stability for both variables, it is particularly compelling to find across-time prediction of ER from EC. These findings demonstrate that these constructs are interrelated at an earlier age than previously found.

However, 18-month EC did not predict 30-month ER. Because EC is just emerging at 18 months (Eisenberg et al., 2010), advances in EC might provide building blocks for emerging ER. ER has been linked to emotional flexibility, in that resilient adults are better able to shift their emotional and physiological reactions in response to changing environmental stimuli (Waugh, Thompson, & Gotlib, 2011). This type of flexibility may depend on the emerging cognitive structures evident in EC. Consistent with this assumption, the lagged paths from ER to EC were nonsignificant, suggesting that EC contributes to individual differences in ER over time rather than vice versa. However, the data are correlational so causal relations cannot be proved.

Parenting also seems to play a role in the emergence of EC and ER. Intrusive parenting was concurrently negatively related to ER at both 18 and 30 months. Moreover, although there were no significant direct paths across time from intrusive parenting to ER, a significant indirect effect in which EC mediated the relation of intrusive parenting to later ER was found. In contrast, although parenting was not related to EC concurrently in the model, both lagged paths to EC were significant. Thus, parenting might affect not only concrete capacities such as being able to control attention and inhibit behavior, but may additionally affect children's abilities to use their skills in an adaptive manner in stressful contexts.

Also of interest, there was no evidence that less regulated children elicited more intrusive parenting and thereby affected ER. However, our sample may not have had the power to provide a strong test of bidirectional across-time paths, especially for weaker paths. Zero-order correlations were nonsignificant between T1 EC and T2 intrusive parenting, but there was a significant correlation between T2 EC and T3 parenting, supporting a weak relation of EC to parenting over time.

The current study has limitations. The sample was not diverse in ethnicity so the results may not be generalizable to other populations. Also, we used a single methodology to assess ER and EC (combined mother and nonparental caregiver reports). However, only about 35% the nonparental caregivers were the same from one assessment to the next, which helps reduce potential bias. Lastly, our model fit for the CFI and TLI were below the recommendations of .95 suggested by Hu and Bentler (1999). However, our findings provide evidence that individual differences in ER are evident in the preschool years and that EC and parenting potentially affect its development.

References

Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Minnesota symposia on child psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.

Block, J., & Block, J. H. (2006). Venturing a 30-year longitudinal study. *American Psychologist, 61*, 315–327. doi:10.1037/0003-066X.61.4.315

Chuang, S. S., Lamb, M. E., & Hwang, C. P. (2006). Personality development from childhood to adolescence: A longitudinal study of ego-control and ego-resiliency in Sweden. *International Journal of Behavioral Development, 30*, 338–343. doi:10.1177/0165025406072795

Cumberland-Li, A., Eisenberg, N., & Reiser, M. (2004). Relations of young children's agreeableness and resiliency to effortful control and impulsivity. *Social Development, 13*, 193–212. doi:10.1111/j.1467-9507.2004.000263.x

Denissen, J. J. A., van Aken, M. A. G., & Dubas, J. S. (2009). It takes two to tango: How parents' and adolescents' personalities link to the quality of their mutual relationship. *Developmental Psychology, 45*, 928–941. doi:10.1037/a0016230

Edwards, C. P., & Liu, W. L. (2002). Parenting toddlers. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 1. Children and parenting* (2nd ed., pp. 45–71). Mahwah, NJ: Erlbaum.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., Murphy, B. C., Maszk, P., Holgren, R., et al. (1996). The relations of regulation and emotionality to problem behavior in elementary school children. *Development and Psychopathology, 8*, 141–162. doi:10.1017/S095457940000701X

Eisenberg, N., Haugen, R., Spinrad, T. L., Hofer, C., Chassin, L., & Zhou, Q. (2010). Relations of temperament to maladjustment and ego-resiliency in at-risk children. *Social Development, 19*, 577–600. doi:10.1111/j.1467-9507.2009.00550.x

Eisenberg, N., Spinrad, T. L., Fabes, R. A., Reiser, M., Cumberland, A., Shepard, S. A., et al. (2004). The relations of effortful control and impulsivity to children's resiliency and adjustment. *Child Development, 75*, 25–46. doi:10.1111/j.1467-8624.2004.00652.x.

Eisenberg, N., Valiente, C., Fabes, R. A., Smith, C. L., Reiser, M., & Shepard, S. A. (2003). The relations of effortful control and ego control to children's resiliency and social functioning. *Developmental Psychology, 39*, 761–776. doi:10.1037/0012-1649.39.4.761

Graziano, P. A., Keane, S. P., & Calkins, S. D. (2010). Maternal behaviour and children's early emotion regulation skills differentially predict development of children's reactive control and later effortful control. *Infant & Child Development, 19*, 333–353. doi:10.1002/icd.670

Hofer, C., Eisenberg, N., & Reiser, M. (2010). The role of socialization, effortful control, and ego-resiliency in French adolescents' social functioning. *Journal of Research on Adolescence, 20*, 555–582. doi:10.1111/j.1532-7795.2010.00650.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi:10.1080/10705519909540118

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review, 30*, 467–480

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items. *Educational and Psychological Measurement, 54*, 757–765. doi:10.1177/0013164494054003022

Martel, M. M., Nigg, J. T., Wong, M. M., Fitzgerald, H. E., Jester, J. M., & Puttler, L. I. (2007). Childhood and adolescent resiliency, regulation, and executive functioning in relation to adolescent problems and competence in a high-risk sample. *Development and Psychopathology, 19*, 541–563. doi:10.1017/S0954579407070265

Martel, M. M., Pierce, L., Nigg, J. T., Jester, J. M., Dams, K., & Puttler, L. I. (2009). Temperament pathways to childhood disruptive behavior and adolescent substance abuse: Testing a cascade model. *Journal of Abnormal Psychology, 37*, 363–373. doi:10.1007/s10802-008-9269-x

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods, 12*, 23–44. doi:10.1037/1082-989X.12.1.23

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Power, T. G. (2004). Stress and coping in childhood: The parents' role. *Parenting: Science and Practice, 4*, 271–317. doi:10.1207/s15327922par0404_1

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development, 29*, 386–401. doi:10.1016/j.infbeh.2006.01.004

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Child Behavior Questionnaire.

Child Development, *72*, 1394–1408. doi:10.1111/1467-8624.00355

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99–166). New York: Wiley

Shonk, S. M., & Cicchetti, D. (2001). Maltreatment, competency deficits, and risk for academic and behavioral maladjustment. *Developmental Psychology*, *37*, 3–17. doi:10.1037/0012-1649.37.1.3

Smeekens, S., Riksen-Walraven, J., & van Bakel, H. (2009). The predictive value of different infant attachment measures for socioemotional development at age 5 years. *Infant Mental Health Journal*, *30*, 366–383. doi:10.1002/imhj.20219

Stams, G.-J. J. M., Juffer, F., Van IJzendoorn, M. H., & Hocksbergen, R. C. (2001). Attachment-based intervention in adoptive families in infancy and children's development at age 7: Two follow-up studies. *British Journal of Developmental Psychology*, *19*, 159–180. doi:10.1348/026151001166010

Sturge-Apple, M. L., Davies, P. T., Martin, M. J., Cicchetti, D., & Hentges, R. F. (2012). An examination of the impact of harsh parenting contexts on children's adaptation within an evolutionary framework. *Developmental Psychology*, *48*, 791–805. doi:10.1037/a0026908

Waugh, C. E., Thompson, R. J., & Gotlib, I. H. (2011). Flexible emotional responsiveness in trait resilience. *Emotion*, *11*, 1059–1067. doi:10.1037/a0021786

Child Development, July/August 2013, Volume 84, Number 4, Pages 1152–1170

Little Evidence That Time in Child Care Causes Externalizing Problems During Early Childhood in Norway

Henrik D. Zachrisson
Norwegian Institute of Public Health and The Norwegian Center for Child Behavioral Development

Eric Dearing
Boston College

Ratib Lekhal
Norwegian Institute of Public Health

Claudio O. Toppelberg
Harvard Medical School and Judge Baker Children's Center

Associations between maternal reports of hours in child care and children's externalizing problems at 18 and 36 months of age were examined in a population-based Norwegian sample (n = 75,271). Within a sociopolitical context of homogenously high-quality child care, there was little evidence that high quantity of care causes externalizing problems. Using conventional approaches to handling selection bias and listwise deletion for substantial attrition in this sample, more hours in care predicted higher problem levels, yet with small effect sizes. The finding, however, was not robust to using multiple imputation for missing values. Moreover, when sibling and individual fixed-effects models for handling selection bias were used, no relation between hours and problems was evident.

The impact of time spent in early child care on social-emotional development is a contentious issue among social scientists and the public alike. Indeed, there remains a lack of empirical consensus on the topic. One review of the cumulative work in the United States concludes that, among child-care predictors of socioemotional problems, quantity of care is the strongest and most consistent, with more hours predicting more problems (Jacob, 2009). However, more recently, researchers have found this prediction to be sporadically evident, but not robust, when using conservative controls for potential selection bias (McCartney et al., 2010).

The scientific and public debate has, to date, relied heavily on child-care studies in the United States and on one longitudinal study in particular—the NICHD Study of Early Child Care and Youth Development (NICHD SECCYD). For this reason, sociopolitical context must be considered; quantity findings in U.S. samples must be interpreted with attention to the lack of regulatory standards at the federal level for child-care quality and professional development (Love et al., 2003), and a relatively early average age of entry into nonparental care due to the relatively short average length of parental leave (UNICEF Innocenti Research Center, 2008). A greater scientific focus on child care in countries with publicly subsidized and regulated child care is the next logical step for the field. In this study, we examined associations between hours in child care and externalizing behavior problems in a large, population-based longitudinal Norwegian study.

Norway: Child Care in Sociopolitical Context

According to UNICEF, Norway meets or exceeds 8 of 10 benchmarks for early childhood service regulation, standards, and quality (examples of benchmarks are: subsidized and regulated child-care services for 25% of children under 3, 1.0% of GDP spent on early childhood services; UNICEF Innocenti Research Center, 2008). The United States, in contrast, meets 3 of the 10 benchmarks. Whereas child care in U.S. policy is generally treated as an unin-

---

The authors acknowledge the Norwegian Mother and Child Cohort Study, which is supported by the Norwegian Ministry of Health, NIH/NIEHS (Grant N01-ES-85433), NIH/NINDS (Grant 1 UO1 NS 047537-01), and the Norwegian Research Council/FUGE (Grant 151918/S10). The authors express their deep appreciation for Stuart Hauser, MD, PhD, a cherished mentor and friend who made this international collaboration possible. Three of the authors (HDZ, ED, and COT) received support from the Norwegian Research Council. The authors acknowledge feedback from Drs. Burchinal and McCartney, and from three anonymous reviewers.

Correspondence concerning this article should be addressed to Henrik D. Zachrisson, Division of Mental Health, Norwegian Institute of Public Health, PO Box 4404, Nydalen, Oslo 0403, Norway. Electronic mail may be sent to henrikdz@ulrik.uio.no.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0005
DOI: 10.1111/cdev.12040

tended or unfortunate consequence of workforce participation among women, child care in the Norwegian corporatist economy is part of a broader family policy to promote maternal workforce participation and employment rights (Mandel & Semyonov, 2005). Norwegian female workforce participation is among the highest in the Organisation for Economic Co-operation and Development (OECD), with 74.4% of women being employed in 2009, compared to 63.4% in the United States, and most Norwegian children are cared for outside of their homes; in 2009, 79% of all 1- to 2-year-olds, and 97% of all 3- to 5-year-olds attended publicly subsidized center care (Statistics Norway, 2011). Nonetheless, there is considerable variability in the number of hours children spend in child care in Norway (Statistics Norway, 2011), making it ideal for examining links between quantity of care and externalizing problems within a sociopolitical context of regulated, near universal and homogenously high-quality care.

In Norway, parents have about 1 year parental leave at nearly full pay (Ministry of Education, 2011). All children have the right to center care or family day care from age 1, which is heavily subsidized, with maximum fees of approximately NOK 2000 (approximately USD 333/month), reduced according to family income. Quality standards (teacher education and child:staff ratios) and curriculum are regulated by law (Ministry of Education, 2010). In center care, adult:child ratios cannot exceed 3:10 for children younger than 3 and 3:19 for those older; at least one of these adults has to be a trained child-care teacher. Family day care is limited to 10 children and adult:child ratios of 1:5 for children older than 3, and lower for younger children; caregivers must receive weekly supervision from a teacher responsible for pedagogical planning (Ministry of Education, 2006). Although standards are currently not entirely met in all centers (Brenna et al., 2010; UNICEF Innocenti Research Center, 2008), quality is relatively high and homogenous (Winsvold & Guldbrandsen, 2009). However, it is noteworthy that, at least in the United States, only modest correlations have been observed between structural indicators of quality and observed quality of caregiver–child interactions (e.g., NICHD Early Child Care Research Network, 2002).

Child-Care Quantity and Externalizing Problems: Theory and Empirical Work

Developmental contexts have an impact on children's behavioral problems. For instance, contexts characterized by chaos, high stress, or including antisocial role models, increase the likelihood of children developing externalizing problems, a domain that includes aggressive, destructive, and hyperactive behaviors (e.g., Hinshaw, 2002). In the early 1980s, following a large increase in women's participation in the labor force, U.S. researchers began documenting associations between hours in nonmaternal care and externalizing problems (Belsky, 1986; Clarke-Stewart & Fein, 1983). In response, theoretical explanations were offered, most based on attachment or social learning theory, yet none have received consistent empirical support (for a review, see McCartney et al., 2010).

A number of studies have continued to document positive associations between hours in child care and externalizing problems in samples of children from the United States (for a review, see NICHD Early Child Care Research Network, 2003). Using child-care provider and teacher reports of externalizing problems, associations with child-care quantity, repeatedly measured throughout early childhood, have been documented in the SECCYD when children were 2 and 4½ years of age, and once they were in kindergarten, but not when children were 3 (NICHD Early Child Care Research Network, 1998, 2003). Associations with maternal reports of externalizing problems were also found, but only in kindergarten. These associations remained evident after adjusting for a host of family and child-care factors, including quality, measured repeatedly. These findings were less consistent when child-care quantity at specific ages was considered. Further analyses indicated associations between more hours in care prior to age 4½, and teacher-reported externalizing behavior up to age 12 (Belsky et al., 2007) and self-reported risk taking and impulsivity at age 15 (Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010).

Evidence from other U.S. studies support the main conclusion from the SECCYD. Higher amounts of nonmaternal care across early childhood have been found to predict negative adjustment reported by teachers and observers at age 5 (Bates et al., 1994) and in third grade (Vandell & Corasaniti, 1990), and maternal reports of behavior problems at ages 3 and 5 (Belsky, 1999). Furthermore, Loeb and colleagues (Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007) found that hours in center care, rather than hours in any nonmaternal care, in the year prior to kindergarten, were associated with poor teacher-rated social skills in kindergarten.

In studies outside of the United States, results have been more mixed. In a recent Australian study, full-time center care (more than 20 hr per week) throughout the child's early years, as well as between ages 2 and 3, was associated with higher levels of parental reports of difficult temperament at age 3, compared to children in part-time care (Yamauchi & Leigh, 2011). In a sample of Dutch twins, children in low and medium amounts of nonparental care were also rated by mothers and fathers as having more externalizing problems at age 3, compared to those exclusively in parental care (van Beijsterveldt, Hudziak, & Boomsma, 2005). Children in high amounts of nonparental care displayed low problem levels similar to those in parental care. Amount of care was defined as proportion of full-time care across the first 3 years.

In a British study, high levels of group care across early childhood were associated with higher levels of teacher-rated antisocial behavior up to age 7, yet the association was not evident at age 10 (Melhuish, 2010). Moreover, in another British sample, no association between average hours in care across early childhood and mother-reported disruptive behaviors was evident at age 3 (Barnes, Leach, Malmberg, Stein, & Sylva, 2009). Similarly, a Canadian study found no association between quantity of nonmaternal care and contemporaneous mother-reported behavior problems in 4- to 5-year-olds (Romano, Kohen, & Findlay, 2010). In Japan, researchers found no association between time spent in "Baby Hotels," child-care centers open 24 hr a day, and observer-rated social competence (Anme & Segal, 2004).

In a Nordic context, two studies have been focused on child-care quantity and child outcomes. More hours in center care at age 3 was associated with higher levels of behavior problems reported by mothers at age 7 in a large Danish sample (Gupta & Simonsen, 2010). In addition, in a Norwegian study using maternal employment across the child's first 4 years as a proxy for early child care, associations were found with teacher-reported problem behavior in 10-year-olds (Borge & Melhuish, 1995), but not with mother reports of behavior problems at ages 4–7.

These previous findings on child-care quantity and externalizing problems vary considerably, as do the designs and measurement across studies. Associations are more consistent in studies using caregiver or teacher reports (e.g., Loeb et al., 2007) than maternal reports (e.g., Barnes et al., 2009) or observer ratings (Anme & Segal, 2004), although associations with maternal reports are found in several studies (e.g., Gupta & Simonsen, 2010; Yamauchi & Leigh, 2011). Furthermore, associations appear largest and most robust when analyses are based on cumulative hours in care across multiple years in infancy and early childhood (e.g., Vandell & Corasaniti, 1990), although they are also found when quantity is measured at a single time point (e.g., Loeb et al., 2007). Notably, varying findings across design features, including reporter of externalizing problems and operationalization of quantity, are consistent with the report by the NICHD Early Child Care Research Network (2003). Associations were also present, although less consistently, when maternal reports and observer ratings of externalizing problems as well as quantity across more limited time periods were considered.

International Variations in Parental Leave and the Average Age of Entry into Care

When considering international studies, variations in sociopolitical context with regard to parental leave policies and average age of entry into care must be noted. In the United States, parents are granted 12-week unpaid parental leave, with 10 states paying temporary disability benefits for 10 weeks (OECD Early Childhood Education & Care Home Page, 2004). This results in a relatively early average age of entry into nonmaternal care in the United States, with approximately 42% of all children entering care in their 1st year of life in 2005 (National Center for Educational Statistics, 2011).

In the Netherlands, where van Beijsterveldt et al. (2005) found some associations between quantity and parent-reported behavior problems, parents are offered 16 weeks of paid leave followed by 6 months of unpaid leave. In Denmark, where Gupta and Simonsen (2010) found associations between quantity of care at age 3 and parent-reported behavior problems at age 7, parents may take about 6 months leave paid at the unemployment benefit rate, and another 6 months at reduced rate. The United Kingdom, from where findings are mixed, offers 6 months maternity leave with almost full pay, followed by 1 year of unpaid leave. This is reflected in the age of entry in the sample for which Barnes and colleagues (Barnes et al., 2009; Sylva, Stein, Leach, Barnes, & Malmberg, 2011) report no association between quantity and parent-reported problems: Almost 50% of children had no nonparental care during their 1st year of life, whereas about 20% entered care from age 6 months for more than 30 hr per week.

It is difficult based on the existing literature to plot a precise pattern of association between policy and whether researchers did or did not detect associations between child-care quantity and externalizing behavior problems, primarily because: (a) of the various ways in which nations structure policies, (b) the limited number of studies outside the United States, and (c) differences in study design and measurement. Nonetheless, it is clear that a 1-year parental leave policy with pay makes Norway an uncommon context to study relations between child-care quantity and behavior problems because this policy results in almost no children entering nonparental care before age 1.

With this in mind, high quantity specifically in the 1st year of life has been associated with later antisocial behavior (Melhuish, 2010) and problem behaviors (van Beijsterveldt et al., 2005). However, entry into child care during the 1st year has been associated with lower levels of maternal reports of physical aggression in children of low-educated mothers (Coté et al., 2007). It is also worth noting that children receiving high quantities of child care during the first years of life are also likely to receive high quantities at later ages, making it difficult to differentiate the effect of entering care at an early age from the cumulative effect of high quantity of care across early childhood.

Potential Selection Effects: Analytical Approaches and Logical Propositions

Existing studies on child-care quantity and children's emotional–behavioral problems are not randomized controlled experiments, but rather correlations estimated within observational designs. One serious concern regarding internal validity is the potential for selection—omitted variables—to bias estimated associations (Duncan, Magnuson, & Ludwig, 2004; Foster, 2010; McCartney, Bub, & Burchinal, 2006). Unmeasured child or family factors may influence both family child-care choices and children's behavior, biasing the estimated effects of child-care quantity per se, by either artificially inflating or deflating the findings.

All of the studies we have discussed, thus far, share an approach to adjusting for potential selection bias: general linear models that include measured child, family, and often also child-care variables as covariates. Even large covariate sets, however, cannot capture all potential sources of selection (Duncan et al., 2004), and improperly specified covariates can bias associations upwardly (Foster, 2010). More conservative approaches to handling selection bias in nonexperimental designs have been recommended, including the study of within-family variation (i.e., sibling fixed-effects models) and within-child variation across time (i.e., individual fixed-effects models; Duncan et al., 2004; McCartney et al., 2006).

Extending these recommendations, McCartney et al. (2010) use logical propositions to test whether associations of child-care quantity and problems are evident in ways consistent with causal hypotheses. In a reexamination of SECCYD data, these authors tested five propositions, four of which move beyond the measured covariate approach used in prior studies. Two of these four tests provided evidence consistent with causality: (a) when conditioning on time spent in child care after age 2, time spent in child care in infancy and toddlerhood was associated with externalizing behavior in kindergarten, and (b) amount of time spent in child care was associated with externalizing behaviors, controlling for earlier problems. However, when estimating within-child change—individual fixed-effects models—the evidence was mixed: Changes in child-care hours between 36 and 54 months predicted changes in externalizing problems, but changes in hours between 24 and 54 months did not. Moreover, in a dose–response analysis, increased hours did not predict increased problems.

In this study, we follow McCartney and colleagues in taking a rigorous data analytic approach toward demonstrating associations that are more or less consistent with the causal hypothesis that more hours in care leads to more problems, extending this methodological practice to a non-U.S. sample. We add to the international evidence on child-care quantity and externalizing problems using a longitudinal, population-based sample in Norway, where care quality is relatively high and homogenous, and the typical age of entry is after 1 year.

The Present Study

In the present study, we examined associations between hours in child care and maternal reports of externalizing problems at 18 and 36 months of age in a sample of 75,271 Norwegian children, 17,910 of whom were siblings. Specifically, we took three general approaches to examining these associations. First, we examined contemporaneous, lagged, and cumulative hours of care as predictors of externalizing problems, using estimates of between-child differences, as has become conventional in the field. By estimating contemporaneous, lagged, and cumulative hours in care as predictors, we allowed for three possible ways in which child-care quantity can predict externalizing problems. Our contemporaneous models assume acute effects of higher hours in care, with rapid onset of externalizing behavior. Our lagged models assume delayed onset of externalizing problems. In turn, our cumulative models (average hours in care across 18 and 36 months) assume additive effects of hours in care, with consequences accumulating across time. Also note, by comparing results from this first set of modeling approaches, we were able to make inferences concerning a dose–response hypothesis. If a higher dosage of child care mattered, then our cumulative models should evidence larger associations than our contemporaneous or lagged models.

Second, we estimated sibling fixed-effects models, examining whether differences between siblings in number of hours in care at the two ages of interest are associated with differences between siblings in levels of externalizing problems. For these sibling fixed-effects models, we also examined contemporaneous, lagged, and cumulative specifications. One advantage of this method (compared to conventional nonexperimental comparisons of children from different families) is that bias caused by omitted variables shared by siblings could be ruled out. In particular, sibling fixed-effects models are useful for ruling out bias caused by persistent, shared parent and family characteristics that are unobserved. Notably, sibling influence constitutes a potential source of bias in these models. It is therefore pertinent to test for sufficient variability in both hours in care and in externalizing problems within sibling pairs.

Third, and finally, we estimated individual fixed-effects models, using within-child variations in hours of care as a predictor of within-child variations in externalizing problems across the two ages of interest. One advantage of this method, compared to conventional nonexperimental comparisons of children, is that all unmeasured sources of bias that are constant over time are eliminated. For example, time-invariant characteristics of children and their parents can be ruled out as potential sources of bias. Note, however, that all techniques for causal modeling in observational data, including sibling and individual fixed-effects models, rely on assumptions that cannot be empirically tested.

For all three of our modeling approaches, we examined (a) both linear and nonlinear associations between hours in care and externalizing problems, with a particular interest in the potential for this association to become increasingly large at increasingly higher levels of hours and (b) estimates before and after imputation of missing data, which, to the best of our knowledge, has been done in only one of the previous studies discussed (Vandell et al., 2010). We interpret our findings within the context of both the strengths and limitations of these data. In particular, we note the value of conservative approaches to selection bias and estimates based on a large and population-based sample from Norway, while keeping in mind that these estimates were based exclusively on maternal report of externalizing problems, only two assessments of child-care quantity and no assessment of quantity prior to 18 months of age (due to the ecology of child care in Norway, i.e., extended parental leave and minimal child-care utilization before age 1), and a sample with considerable attrition.

Method

Participants

Data from the population-based Norwegian Mother and Child Cohort Study (MoBa; for a complete description, see Magnus et al., 2006, and www.fhi.no/moba-en) were used in this study. All women in Norway giving birth between late 1999 and 2010 at hospitals and maternity units with more than 100 births annually, altogether 52 units, were eligible for the study—there are no exclusion criteria for participation. Women were invited to participate when they attended routine ultrasound examinations offered to all pregnant women in Norway at the 17th week of gestation. Information on health, lifestyle, and child development was collected by questionnaire during pregnancy at the 17th, 22nd, and 30th weeks of gestation and after birth by mail when the child was 6, 18, and 36 months of age.

As of October 2010, 90,725 mothers of 108,639 children had enrolled and completed baseline assessments, which represented 42.1% of all eligible mothers in Norway. Of the eligible children whose mothers enrolled, 69.3% ($n = 75{,}271$, including 17,910 siblings) were born by October 2007, making them eligible for inclusion in the present analyses because they were old enough for mothers to complete the 6-, 18-, and 36-month questionnaires. Among these children eligible for analyses, maternal questionnaire response rates at 18 and 36 months (the ages when hours in child care were assessed) were 72.4% and 59.3%, respectively.

Potential self-selection bias in the MoBa was examined by means of differences in prevalence estimates and association measures between MoBa participants and all women giving birth in Norway on demographics, health-related behaviors, and on a number of pregnancy- and birth-related variables (Nilsen et al., 2009). There were some differences between MoBa participants and nonparticipants. Young mothers (less than 25 years) and those living alone are underrepresented relative to the Norwegian population. Moreover, mothers participating have fewer health-related risks like smoking, and their children have higher birth weights and in general better neonatal health (including Apgar score) than children of those not participating. However, the relative differences were small (0.3%–1.2%). Importantly, Nilsen et al. (2009) have demonstrated that despite risk prevalence differences between the MoBa sample and the population, associations between risk exposures and child development outcomes available through public registries are not statistically different when MoBa participants are compared with the population of Norway.

Measures

Externalizing Problems

These were measured at 18 and 36 months by using selected items (9 items at 18 months and 11 items at 36 months were included in the MoBa) from the mother-reported Child Behavior Checklist for Ages 2-3 (CBCL/2-3; Achenbach, 1992). Items were selected by a team of four clinical and developmental psychologists, based on clinical and theoretical standards, as well as empirical representativeness (high factor loadings) for externalizing behavior. Mothers rated whether each item statement reflected their child's behavior during the last 2 months from 1 = *not true* to 3 = *very true or often true*.

At 18 months, 3 items from the Attention Problems scale (of a total of 5 items) were used: "Punishment doesn't change his/her behavior," "Can't sit still, restless or hyperactive," and "Quickly shifts from one activity to another." In addition, at 18 months, 6 items from the Aggressive Behavior scale (of 20 items) were used: "Defiant," "Doesn't seem to feel guilty after misbehaving," "Gets into many fights," "Gets into everything," "Hits others," "Punishment doesn't change his/her behavior." At 36 months, 1 item was added from the Attention Problems scale: "Poorly coordinated or clumsy," and 1 item was added from the Aggressive Behavior scale: "Demands must be met immediately." (Note that for our individual fixed-effects models, we used a shorter scale at 36 months, based on the 9 items used at both assessments.) Scale reliability was adequate at both time points (Cronbach's αs = .62 at 18 months and .74 at 36 months), with confirmatory factor analyses indicating adequate fit at 36 months (comparative fit index [CFI]/Tucker–Lewis index [TLI] = .866/.832, root mean square error of approximation [RMSEA] = .086), but somewhat less adequate fit at 18 months (CFI/TLI = .798/.731, RMSEA = .092). We examined whether the subset of the 11 items used at 36 months was representative of the full externalizing broad band scale of the CBCL in the NICHD SECCYD data, and found a correlation of .92. In accordance with recommendations by Achenbach (1992), for when a selection of items from the CBCL (rather than the complete scale) is used, we report raw scores rather than *T* scores.

As a complement to the CBCL, four items from the Infant-Toddler Social and Emotional Assessment (ITSEA; Carter, Briggs-Gowan, Jones, & Little, 2003) were also used to assess child social-emotional problems at 36 months. See online supporting information Appendix S1.

Child Care

The number of hours per week that children spent in nonparental child care was reported by the mothers when their children were 18 and 36 months of age. At these time points, mothers also reported on the type of child-care arrangement (home care by mother or father, unqualified child minder, family day care or outdoor nursery, or center care) that represented the child's primary care arrangement. At 18 months, mothers additionally reported retrospectively which of these types of care the child had experienced from 0 to 18 months of age (in the intervals 0–6, 7–9, 10–12, 13–15, and 16–18 months).

Risk Factors for Externalizing Problems

Family and child risk factors were considered as covariates for relations between child-care hours and externalizing problems.

Family and prenatal risk factors. Maternal and paternal education, partner status (single vs. partnered), non-Norwegian family background, and family income (adjusted for consumer price index with 2005 as reference year) were reported by the mothers at 17th gestational week. Perceptions of economic hardship were assessed at 6 and 18 months after birth, using Likert-type responses to the question "Have you found it difficult sometimes during the last 6 months to cope with running expenses for food, transport, rent, etc.?"

Mothers also reported on their psychological distress (anxiety and depression) using the Hopkins Symptom Checklist (SCL; Hesbacher, Rickels, Morris, Newman, & Rosenfeld, 1980), at 17th gestational week and when their child was 6, 18, and 36 months old. Parental locus of control was measured at 36 months with five questions from the Parental Locus of Control scale (Campis, Lyman, & Prentice-dunn, 1986) and four supplementary questions developed for the MoBa. In addition, adverse life events were measures at 18 and 36 months by a checklist of 11 events, including "problems at work" and "have you had problems or conflicts with your family, friends, or neighbors." Medical Birth Registries' information on number of cigarettes smoked per week during pregnancy was included.

Child risk factors. Medical Birth Registries' information on child gender, multiple births, birth weight (dichotomized; < 2500 and > 2500 g), Apgar scores 5 min after birth, and congenital syndromes (including Down syndrome, cleft lip and palate, and limb malformations) was retrieved.

Statistical Analysis

Analytic Approach 1: Between-Child Ordinary Least Squares (OLS) Models

As a first analytical approach we estimated associations between contemporaneous, lagged, and cumulative hours in child care and externalizing problems in a series of OLS regression models, conditioned on the set of 30 child- and family-risk covariates listed in Table 1. For example, the following equation is a summary of the conditional models in which the linear association between hours in care and externalizing problems is estimated while controlling for study covariates: $\text{Ext}_i = \beta_{00} + \beta_{01}(\text{CCHours}_i) + \beta_{02}(\text{CovX}_i) + \ldots + u_i$. We also estimated nonlinear associations in these models by including a quadratic term for child-care hours.

In the contemporaneous models, externalizing behavior problems at 18 months were regressed on child-care hours at 18 months and, similarly, problems at 36 months were regressed on hours at 36 months. In the lagged models, externalizing behavior problems at 36 months were regressed on child-care hours at 18 months, conditioning on child-care hours at 36 months. In the cumulative models, externalizing behavior problems at 36 months were regressed on children's cumulative hours in care at 18 and 36 months, operationalized as their mean level of hours across these two time points. In the contemporaneous model for externalizing problems at 18 months, we included only covariates that were measured at 18 months or earlier. In all other models, we included the full covariate set.

Analytic Approach 2: Sibling Fixed-Effects Models

For the more than 17,000 siblings in the sample, we estimated within-family associations between hours in child care and externalizing problems as sibling fixed effects. Consider, for example, the following equation for linear hours in care (simplified in this example to include only sibling pairs, ignoring covariates and the error term): $\text{Ext}_{\text{sib1}} - \text{Ext}_{\text{sib2}} = \beta_{10}](\text{CCHours}_{\text{sib1}} - \text{CCHours}_{\text{sib2}})$. In this model, $\text{CCHours}_{\text{sib1}}$ is the value of child-care hours for the first sibling and $\text{CCHours}_{\text{sib2}}$ is the value of child-care hours for the second sibling. As such, β_{10} should be interpreted as the average within-family association between hours and externalizing problems. Our sibling fixed-effects models parallel the OLS between-child models in that we estimated contemporaneous, lagged, and cumulative hours in care as predictors of externalizing problems. In addition, we estimated both linear and nonlinear specifications of hours, in the latter case adding a quadratic child-care term to the model. These sibling fixed-effects models controlled for potential bias by unobserved variables that were fixed within families (i.e., constant across siblings) such as shared family environments. In addition, we included 30 covariates that could differ across siblings.

Analytic Approach 3: Individual Fixed-Effects Models

Within-child associations between changes in hours in child care and changes in externalizing problems (between 18 and 36 months) were estimated in individual fixed-effects models. Consider, for example, the following model: $\text{Ext}_{it1} - \text{Ext}_{it2} = \beta_{10}(\text{CCHours}_{it1} - \text{CCHours}_{it2})$. In this model (ignoring covariates and error term), CCHours_{it1} is the quantity of hours for child *i* at 18 months and CCHours_{it2} is this child's quantity of hours at time 2. As such, β_{10} should be interpreted as the average within-person association between child-care hours and externalizing problems. As with our other two modeling approaches, we estimated both linear and nonlinear specifications of hours, in the latter case adding a quadratic child-care term to the model. Compared to our other modeling approaches, these individual fixed-effects models

Table 1
Summary of Means, Standard Deviations, and Percentage of Complete Data Coverage (N = *75,271)*

Variable	*M* (*SD*)/%	Range	% covered
Child-care quantity			
Hours in nonmaternal care 18 months	24.25 (14.44)	0–80	65.1
Hours in nonmaternal care 36 months	29.45 (11.11)	0–99	53.2
Externalizing behavior			
18 months 9 items	1.48 (0.28)	1–3	72.9
36 months 11 items	1.51 (0.29)	1–3	58.3
36 months 9 items	1.49 (0.30)	1–3	58.3
Child-care variables			
Entry			
< 7 months	0.1%		72.3
7–9 months	1.1%		72.4
10–12 months	11.6%		72.5
13–15 months	51.7%		72.6
Type of care at 18 months			
Home with mother or father	27.2%		72.5
Unqualified child minder	5.8%		72.5
Family day care	19.9%		72.5
Center care	47.1%		72.5
Type of care at 36 months			
Home with mother or father	5.2%		57.2
Unqualified child minder	0.9%		57.2
Family day care	5.1%		57.2
Center care	88.7%		57.2
Child risk factors			
Boys	51.1%		100
Twins or triplets	3.5%		100
Congenital syndromes	4.98%		100
Serious malformation at birth	2.9%		100
Birth weight < 2500 g	4.3%		100
Apgar 5 min	9.41 (0.79)	0–10	100
Family risk factors			
Maternal education (years)	14.58 (2.50)	8–18	94.1
Paternal education (years)	14.02 (2.74)	8–18	89.9
Family income (1,000 NOK)	594 (196)	0–1144	87.6
Non-Norwegian family background	10.3%		94.4
Mother single at 17th gestational week	2.0%		95.9
Mother single at age 6 months	2.0%		83.1
Mother single at age 18 months	3.5%		71.9
Mother single at age 36 months	3.1%		56.7
Other children in the family	1.34 (0.77)	0–8	65.2
Cigarettes smoked weekly during pregnancy	2.11 (8.43)	0–268	100
Economic hardship 6 months	1.28 (0.62)	1–4	83.7
Economic hardship 18 months	1.33 (0.65)	1–4	68.0
Maternal distress 17th gestational week	1.27 (0.40)	1–4	95.2
Maternal distress at age 6 months	1.24 (0.35)	1–4	84.1
Maternal distress at age 18 months	1.28 (0.37)	1–4	70.9
Maternal distress at age 36 months	1.28 (0.39)	1–4	56.5
Parental locus of control	2.34 (0.76)	1–5	54.9
Parental love	4.25 (0.50)	1–5	54.9
Parental inefficacy	1.32 (0.61)	1–5	54.9
Adverse events 18 months	0.82 (1.26)	0–11	70.9
Adverse events 36 months	0.33 (1.38)	0–11	56.5

controlled for potential bias caused by time-invariant unobserved variables—unmeasured child and family characteristics, or other features of developmental context, that were constant over time. These models were also conditioned on all time-varying covariates.

Missing data. In Table 1, we present descriptives for study variables, along with the percentage of complete, nonmissing values. The percentage of missing data due to item nonresponse was less than 2% across all items, with only two exceptions: externalizing behavior items at 18 months (6.5%) and adverse events at 18 and 36 months (14% and 22%, respectively). We replaced missing items in scales with the scale mean.

Missing data due to attrition, however, were more considerable. In total, 65.0% of children (n = 49,000) had complete data on both child care and externalizing problems at 18 months, and 52.8% of children (n = 39,807) at 36 months, and 44% (n = 33,092) across 18 and 36 months. Across all variables and all time points, 78.4% of the data was complete, including 70.3% at 18 months and 55.8% at 36 months. Statistically significant differences between respondents and nonrespondents were evident, but the effect sizes ranged from small to very small. At the 17th gestational week, for instance, those not dropping out by 36 months had higher family income (d = .14), more educated mothers (d = .07), and less depressed and anxious mothers (d = .12); at 18 months, those dropping out at 36 months spent, on average, more hours in nonparental care (d = .06) and had higher levels of externalizing problems (d = .02).

Given these missing data, we took two primary approaches to our analyses. First, we estimated models using listwise deletion (i.e., only children with complete data on all study variables were included in the analyses, and children with one or more missing values were excluded). Nonoverlapping patterns of missing values across child care, externalizing problems, and covariates led to estimation sample sizes ranging from 24,804 to 35,831 children for the between-child OLS models, 2,042 to 2,796 families in the sibling fixed-effects models, and 31,150 children in the individual fixed-effects model.

Second, following best-practice recommendations for handling moderate to large amounts of missing data, we used multiple imputation (MI; Graham, 2009). We estimated 20 data sets based on all covariates in Table 1, including quadratic terms of hours in care, using PASW 17.0 (SPSS Inc., 2009), with fully conditional specification of the multivariate model by a series of conditional linear models, one for each incomplete variable (van Buuren, 2007).

Results

Descriptive Statistics and Preliminary Analyses

Three points based on the descriptive data in Table 1 are worth noting. First, on average, at 18 months compared to 36 months, children were more likely to be exclusively in parental care (27.2% vs. 5.2%) and, on average, were in fewer hours of nonparental care (24.25 hr vs. 29.45 hr). Nonetheless, there was substantial between-child variability at both time points; at 36 months, for example, children with 1 *SD* below the mean were in nonparental child care for approximately 18 hr per week, and children with 1 *SD* above the mean were in nonparental care for about 40 hr.

Second, although children in nonparental care were distributed across settings, most were in center care (47.1% at 18 months and 88.7% at 36 months), with smaller percentages in family day care (19.9% and 5.1%, respectively), and even fewer with unqualified child minders (5.8% and 0.9%, respectively). Third, as displayed in online supporting information Table S1, hours in care and externalizing problems demonstrated notable instability over time, at least with regard to children's rank order. Although hours of care at 18 and 36 months were moderately correlated (r = .37 for the nearly 40,000 children with complete data at both time points), over 85% of the variance was nonoverlapping across time. Similarly, for children's levels of externalizing problems across time, the association was moderate in size (r = .45 for children with complete data), regardless of whether 9 or 11 items were used at 36 months.

There were no detectable associations between hours in care and externalizing problems at either 18 or 36 months (see online supporting information Table S1); all of the correlation coefficients were nearly zero. However, these correlations assume a linear relation. As such, we also fit fractional polynomials to the data in an effort to determine whether the best fitting curve was linear or nonlinear. In particular, we were concerned that the association between hours and problems could be increasingly stronger at increasingly higher levels of hours.

As can be seen in online supporting information Figure S1, the association between hours and externalizing behavior problems took a quadratic form;

at both 18 and 36 months, there was little association between hours and externalizing problems at the low end of the hours' distribution, but an increasingly large association at the high end of the distribution. The strength of this nonlinear association appeared greater prior to using multiple imputation, but the effect size was relatively small in either case at 36 months. Using listwise deletion at 36 months, for example, the estimated difference in externalizing problems between children in 45 hr of care—approximately 1.5 *SD* above the mean—and those in 30 or fewer hr of care was less than 15% of 1 *SD*. The effect size was larger at 18 months, at least at the highest end of the distribution for hr when using listwise deletion; however, the estimated difference in problems between children in 45 hr of care and those in 30 or fewer hr was still less than 20% of 1 *SD*.

Two points regarding these unconditional patterns are worth noting, however. First, less than 4% of children were in nonparental care for more than 40 hr per week at 18 or 36 months, and less than 1% were in nonparental care for more than 50 hr. Second, the unconditional estimates were primarily for purposes of determining the shape of association; they were not adjusted for potential selection effects.

Estimating the Consequences of Contemporaneous, Lagged, and Cumulative Hours

Following up on our preliminary estimates of the shape of association, we more rigorously examined relations between hours in child care and child problems using conditional between-child OLS models in which we statistically controlled for the set of family and child risk factors detailed in Table 1. Specifically, we estimated contemporaneous, lagged, and cumulative measures of hours in care in these conditional OLS models. In the contemporaneous models, externalizing problems at 18 months were regressed on hours at 18 months, and a parallel model was estimated for these variables at 36 months. Problems at 36 months were regressed on hours at 18 months in the lagged models and on average hours at 18 and 36 months in the cumulative models. For these analyses, we first used listwise deletion for missing values and then estimated our models using 20 multiple imputation data sets. (We also estimated our models using 5 and 100 multiple imputation data sets, with substantively identical results to those presented here.) For all of our models, we estimated two specifications of hours in child care: (a) linear and (b) a nonlinear specifications (i.e., the nonlinear specification included both a linear and quadratic term for hours in the equation). An overview of results from these OLS models is provided in Table 2; note that coefficients for hours in care correspond to 10-hr increments of child care and are presented for raw and standardized externalizing scores.

There were statistically significant linear associations between hours and child problems in all three models using listwise deletion (i.e., contemporaneous, lagged, and cumulative specifications) and two of the three models using multiple imputation (i.e., contemporaneous and cumulative specifications). The linear effect sizes were, however, very small such that a 10-hr difference in hours was associated with 4% of 1 *SD* difference in externalizing problems (i.e., .01 points) or less. However, as expected based on our fitted fractional polynomial curves, hours in care also demonstrated nonlinear relations with externalizing problems, at least when using listwise deletion for missing values.

In Figure 1, based on estimates from both the listwise deletion and multiple imputation models, we display the nonlinear association at 18 months as well as the nonlinear contemporaneous, lagged, and cumulative associations at 36 months. In all of the listwise deletion models, there was a slight decrease in externalizing problems when comparing children in no child care with those in increasing hours of care, up to approximately 20 hr. Beginning at about 20 hr of care, however, a higher quantity of care was associated with an increasingly higher level of externalizing problems in the listwise deletion models. This nonlinearity was most pronounced for the contemporaneous estimate at 36 months, although children in 45 hr of care displayed problem levels only 15% of 1 *SD* higher than those in 20 hr of care. Note that few children spend more than 40 hr in care.

In addition, as can also be seen in Figure 1, none of these nonlinear findings proved to be robust when we estimated using multiple imputation for missing values. Although estimate precision (i.e., the standard errors) was fairly similar or better in multiple imputation models compared to listwise deletion models, the quadratic coefficients were, at most, only 15% as large in multiple imputation models.

Sibling Fixed Effects: Within-Family Estimates

We also estimated sibling fixed effects for associations between hours in child care and externalizing problems taking advantage of the 17,910

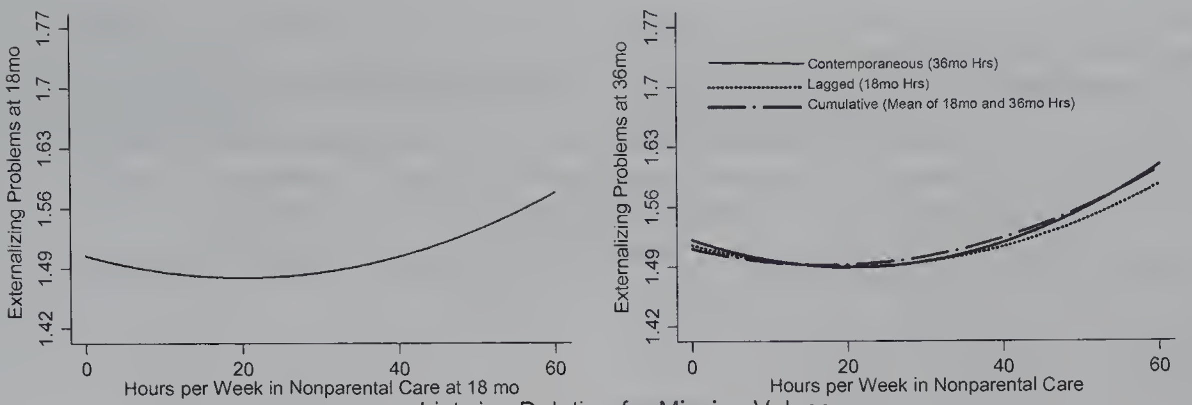
Table 2

Summary of OLS Regression of Hours in Child Care as Predictors of Child Externalizing Problems

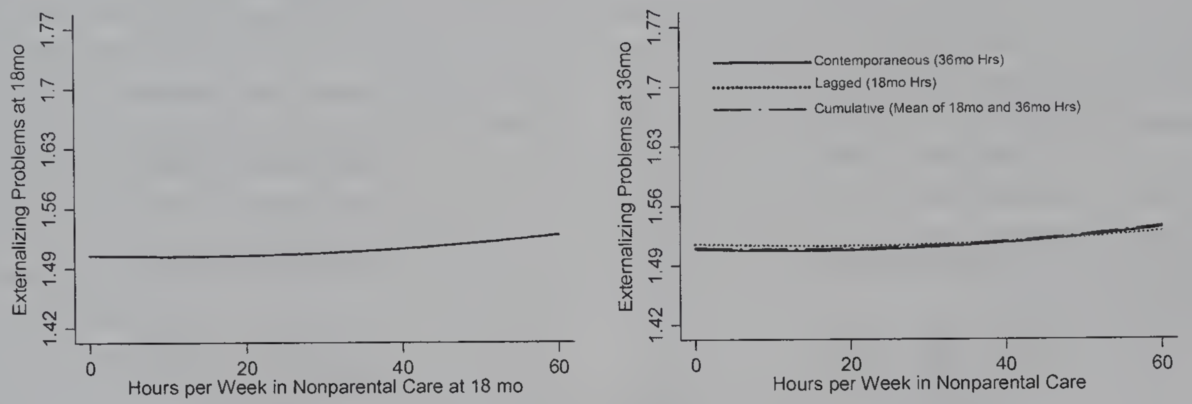
	Listwise-deletion models			Multiple-imputation models ($n = 75,271$)		
	Contemporaneous B (*SE*) [STDY]	Lagged B (*SE*) [STDY]	Cumulative B (*SE*) [STDY]	Contemporaneous B (*SE*) [STDY]	Lagged B (*SE*) [STDY]	Cumulative B (*SE*) [STDY]
18-month CBCL	$n = 35,831$					
Linear estimate	.005** (.001) [.016]			.001 (.001) [.004]		
Nonlinear estimate						
Hours	−.025*** (.005) [−.089]			−.002 (.002) [−.006]		
Hours^2	.006*** (.001) [.022]			.001* (.001) [.003]		
36-month CBCL	$n = 28,822$	$n = 24,804$	$n = 26,290$			
Linear estimate	.007*** (.002) [.024]	.004* (.002) [.014]	.010*** (.006) [.035]	.005** (.001) [.013]	.002 (.001) [.006]	.004* (.001) [.011]
Nonlinear estimate						
Hours	−.032*** (.007) [−.109]	−.025*** (.007) [−.086]	−.022** (.008) [−.074]	−.003 (.006) [−.008]	−.003 (.004) [−.008]	−.003 (.008) [−.008]
Hours^2	.008*** (.001) [.026]	.006*** (.001) [.021]	.006*** (.002) [.021]	.001 (.001) [.004]	.001 (.001) [.003]	.001 (.001) [.003]

Note. All multiple-imputation models are based on 20 imputed data sets. Covariates at 18 months: type of child care, parental education, family income, non-Norwegian background, number of siblings, single parent (average to 18 months), smoking during pregnancy, economic hardship (average 6 and 18 months), maternal depression (average to 18 months), adverse events at 18, gender, multiple births, congenital syndromes and serious malformation, low birth weight, Apgar score, and year of birth. Covariates at 36 months, the lagged models, and the cumulative models: type of child care at 18 and 36 months, parental education, family income, non-Norwegian background, child single parent (average to 36 months), number of children in the family, smoking during pregnancy, economic hardship (average 6 and 18 months), maternal depression (average to 36 months), adverse events (average 18 and 36 months), parenting (36 months), gender, multiple births, congenital syndromes and serious malformation, low birth weight, Apgar score, and year of birth. OLS = ordinary least squares; STDY = standardized coefficients (change in standard deviation units of problems given 10 hr increase); CBCL = Child Behavior Checklist.

$^*p < .05$. $^{**}p < .01$ $^{***}p < .001$.



Listwise Deletion for Missing Values



Multiple Imputation for Missing Values

Figure 1. Associations between hours in child care and externalizing problems.
Note. Between-child ordinary least squares (OLS) estimates of the conditional nonlinear associations between hours in child care and externalizing problems at 18 and 36 months, using listwise deletion (top two graphs) and multiple imputation for missing values (bottom two graphs). The range on the *y*-axis represents approximately 1 *SD* on externalizing problems, and the distance between horizontal bars originating on this axis is approximately 20% of 1 *SD*.

siblings in the data set (of these 2,627 are twins or triplets). That is, we estimated within-family associations, examining whether differences between siblings' hours in child care predicted differences in their levels of externalizing problems. These models closely paralleled our other models in that we examined contemporaneous, lagged, and cumulative hours in care. And we estimated these models using listwise deletion and multiple imputation for missing data.

The sibling models were conditioned on twin or triplet status, number of siblings in the family, as well as on all child and family background factors that varied across siblings (i.e., a total of 30 child and family covariates were included). It is worth noting that one assumption of sibling fixed-effects models is that there is variability across siblings, in this case with regard to hours in care and externalizing problems. In this study, there were considerable differences in the quantity of child care experienced by siblings; at 18 and 36 months, standard deviations for within-family differences in hours in care were 10.34 and 9.48 hr, respectively, and for externalizing problems 0.22 and 0.22, respectively.

A summary of our sibling fixed-effects results are presented in Table 3. In short, the estimated effects of hours in child care were null in all of the sibling models; levels of externalizing behavior problems for siblings in more or less hours of care were statistically indistinguishable from one another. Beyond statistical significance, the direction of association between hours and externalizing problems varied across specifications—compare the contemporaneous and lagged models with the cumulative models—and effect sizes were very small (i.e., a 10-hr difference in child care was associated with less than 5% of 1 *SD* difference in externalizing behavior problems between siblings).

Individual Fixed Effects: Within-Child Estimates of Change

In addition to sibling fixed effects, we estimated individual fixed effects for the association between hours in care and externalizing problems. In other words, we examined whether within-child changes in hours predicted within-child changes in externalizing problems. These models were conditioned on time-varying covariates (i.e., a total of seven

Table 3
Summary of Sibling and Individual Fixed-Effects Models

	Listwise-deletion models		Multiple-imputation models	
	Sibling fixed-effects B (*SE*) [STDY]	Individual fixed-effects B (*SE*) [STDY]	Sibling fixed-effects b (*SE*) [STDY]	Individual fixed-effects B (*SE*) [STDY]
18-month CBCL	*n* = 5,592			
Linear estimate	.007 (.005) [.025]		.001 (.003) [.004]	
Nonlinear estimate				
Hours	.012 (.020) [.042]		−.002 (005) [−.004]	
Hours^2	−.001 (.004) [−.004]		.001 (.001) [.003]	
36-month CBCL	*n* = 4,084			
Linear estimate	−.003 (.007) [−.011]		.001 (.004) [.002]	
Nonlinear estimate				
Hours	.015 (.025) [.050]		.008 (.016) [.024]	
Hours^2	−.004 (.004) [−.013]		−.001 (.003) [−.004]	
Lagged				
Linear estimate	−.002 (.013) [−.007]		−.001 (.004) [−.003]	
Nonlinear estimate				
Hours	.006 (.042) [.021]		−.002 (.005) [−.005]	
Hours^2	−.002 (.009) [−.006]		.000 (.000) [.000]	
18–36-month CBCL	Cumulative	Within-child change *n* = 31,150	Cumulative	Within-child change
Linear estimate	−.008 (.015) [−.026]	−.001 (.001) [−.003]	−.001 (.004) [−.003]	−.001 (.001) [−.002]
Nonlinear estimate				
Hours	−.016 (.056) [−.054]	−.006 (.005) [−.020]	−.001 (.005) [−.003]	−.000 (.005) [−.001]
Hours^2	.002 (.011) [−.006]	.001 (.001) [.004]	.000 (.001) [.000]	−.000 (.001) [−.000]

Note. Multiple-imputation models were based on 20 imputed data sets. Sibling fixed effects were restricted to families with siblings and modeled sibling differences in outcomes as a function of sibling differences in hours. Covariates for the sibling fixed-effects models at 18 and 36 months were the same as those included in between-child OLS analyses. Individual fixed effects modeled differences in outcomes over time as a function of differences in hours over time; covariates included changes from 18 to 36 months in: single parent, adverse events, maternal depression, and type of child care. OLS = ordinary least squares; STDY = standardized coefficients (change in standard deviation units of problems given 10 hr increase); CBCL = Child Behavior Checklist.

child and family factors that varied over time), and we examined both linear and nonlinear specifications for hours in child care. And, again, we estimated both listwise deletion and multiple imputation specifications. It is also worth noting that unconditional individual fixed effects in this study demonstrated considerable within-person change in hours of care, with 1 *SD* of 11.3 hr from 18 to 36 months.

In short, we found no evidence that changes in hours of care predicted changes in externalizing problems (see Table 3), despite substantial statistical power even in the listwise deletion models; indeed, in the individual fixed-effects models using multiple imputation for missing data, the coefficients for hours in care were all very close to zero in absolute value.

As a final step in our analytic plan we conducted robustness checks and examined potential moderators of associations between child-care hours and behavior problems. As can be seen in online supporting information Appendix S1, the robustness checks did not reveal substantive differences from the analyses presented here, and we did not find child gender or age of entry to moderate the association between quantity and externalizing problems.

Discussion

Debate over the prospect that time spent in early child care may increase children's risk of developing externalizing behavior problems has led researchers to call for two advances in the field. First, there has been a call for more conservative approaches to controlling for potential selection effects (McCartney et al., 2010). Second, the value of increased study of child care in countries that have progressive child-care policies has been emphasized (Love et al., 2003). In response to these calls, the purpose of the current study was to examine associations between hours in child care and externalizing behavior problems during early childhood for a large sample of Norwegian children.

We applied a variety of statistical approaches with the dual goals of: (a) taking increasingly conservative steps toward control for potential selection bias and (b) covering an array of theoretical forms that relations between hours in care and externalizing problems might take if, in fact, such relations were causal. With considerable missing data, we estimated all models using listwise deletion and multiple imputation. Using conventional covariate adjustment for selection bias, effect sizes from models based on listwise deletion were relatively larger than those from models based on multiple imputation, albeit small in either case. When we used more conservative adjustments for selection bias, associations between child-care hours and externalizing problems were consistently very close to zero and null, regardless of approach to missing data. Given Norway's comprehensive early childhood policy focused extended paid parental leave through the child's 1st year of life and on near universal high-quality care, these results contribute to the cumulative knowledge on child care in a sociopolitical context of quality standards and oversight.

Regardless of modeling strategy, there was very little evidence that hours in child care had much of an influence on externalizing problems at levels of 40 hr or less. Across the lower half of the distribution on hours, from those children who were exclusively in parental care through those who were in approximately 40 hr of nonparental child care, there were only very small differences in externalizing problems, differences that were statistically significant in only one of our models. In some models, however, especially those based on listwise deletion, we did observe somewhat larger externalizing problem differences at increments of greater than 40 hr. However, even in our least conservative models using listwise deletion, less than 20% of 1 *SD* in externalizing problems separated children in fewer than 30 hr of child care from those in 45 hr of care. Beyond small effect sizes, the practical significance of finding slightly elevated problem levels among those in the most hours of care was further limited by the fact that less than 4% of children in this sample were in child care for more than 40 hr.

Moreover, it is worth noting that differences in externalizing behaviors of up to 20% of 1 *SD* are very far from what is considered "clinical" levels of externalizing behavior problems. These are typically characterized by elevations above 1.5 or 2 *SD* (corresponding to "borderline" or "clinical" levels; Achenbach, 1992). Thus, even in our least conservative models, the differences in levels of externalizing behavior problems between children in very few hours of care and those in 45 or more hr of care were variations within the normal range, rather than pathological manifestations.

One of the most important findings in this study was the lack of evidence from our sibling or individual fixed-effects analyses that variations in hours—either within families or within children, over time—predicted externalizing problems, regardless of whether these models adjusted for attrition using multiple imputation. or not. Given considerable statistical power in these models and considerable variability in hours of child care across siblings and across time for individual children, these null results were not consistent with the causal hypothesis that more hours in child care leads to more externalizing problems. None of our modeling approaches provided a panacea for the internal validity questions inherent in nonexperimental work, but one potential explanation of the null results is that the sibling models held constant any persistent unobserved factors within families and the individual fixed-effects models held constant any persistent unobserved factors across time that were otherwise upwardly biasing estimates in our OLS models. Interestingly, in a U.S. sample, Jaffe and colleagues also found that associations between age of entry into nonparental care and externalizing problems evident in between-child regression models disappear when more conservative sibling fixed-effects models are applied (Jaffee, Van Hulle, & Rodgers, 2011).

Our current study, juxtaposed with the works of McCartney et al. (2010) and Jaffee et al. (2011) should caution child-care researchers against drawing definite conclusions from between-child regression models that employ covariate adjustment to control for selection (also see Yamauchi & Leigh, 2011, for an argument on the extent to which covariate adjustment underestimates selection effects in child-care research). This is further underscored by a study on preschool center attendance and externalizing behavior of children from low-income families (Crosby, Dowsett, Gennetian, & Huston, 2010). Using covariate-adjusted OLS models, these authors found that center attendance was associated with a small increase in externalizing problems. In contrast, when more conservative control for selection was used—instrumental variable estimates—these authors find the opposite: Center attendance was associated with fewer problems.

A second methodological issue worth considering for this study, and the field, is the handling of missing data. Although effect sizes were always quite small, significant associations between hours in child care and externalizing problems were most often evident prior to using multiple imputation. Perhaps because best-practice recommendations for handling missing data have only recently become well known, we found only two studies in the extant literature that approached missing data with methods other than listwise deletion, namely, Vandell et al. (2010) using full information maximum likelihood estimation, and McCartney et al. (2010) using a dummy variable approach to missing data (M. Burchinal, personal communication, May 7, 2012). To the best of our knowledge, all other reports from the SECCYD, for example, have employed listwise deletion when investigating links between child-care quantity and externalizing problems. Attrition, however, has not been as high in the SECCYD as it was in the study reported here.

Attrition in the current data set was considerable (more than 45% by 36 months), but it was similar to what has been observed in other population-based cohort studies of this size (Szklo, 1998). It appears that children whose mothers dropped out of the study were those likely to exhibit relatively high levels of externalizing problems at low doses of child care and those likely to exhibit relatively low externalizing problems at high doses of child care. In light of our findings, it is worth considering whether selective attrition in child-care studies may upwardly bias associations between hours and externalizing problems.

Beyond analytic approaches to controlling potential selection bias, another pattern evident in our results is worth noting with regard to the causal hypothesis. Our models for cumulative hours in care resulted in somewhat smaller effect sizes than did our models for contemporaneous hours in care. These results seem inconsistent with the causal hypothesis; if hours in child care influences children's externalizing behavior, then one might expect a dose–response relation, with higher total "dosages" of child-care hours resulting in higher externalizing behavior "responses." It is possible that hours in child care influence externalizing problems in an acute manner best detected in the contemporaneous models, but our results differ from other researchers (e.g., for a review, see NICHD Early Child Care Research Network, 2003) who have reported particularly robust associations for cumulative hours. Contemporaneous associations may have been strongest in this study because they were most susceptible to simultaneity bias (i.e., externalizing problems influencing hours in child care). Regardless, beyond statistical issues, we want to emphasize the potential role of sociopolitical context when comparing our findings with those in the United States.

Child Care in Norwegian Context

The sociopolitical context of Norway is considerably different from the United States as well as most other countries in which associations between child-care quantity and child socioemotional outcomes have been examined (e.g., Canada), particularly with regard to child and family policy. Consider, for example, three Norwegian policies with potential implications for child outcomes and the results of this study: near universal access to center care, regulated child-care quality standards, and parental leave policies.

Norway provides near universal access to center care from age 1 year. The goal of universal access has the explicit purpose of reducing social selection into regulated child care (Ministry of Education, 2007). This policy leads most parents to choose center care for their children, yet age of entry is varying. This is illustrated by the fact that in this study, 67% of 18-month-olds and nearly 94% of 36-month-olds attend center care or family day care. In contrast, for instance, in the SECCYD, 38% of the 36-month-olds attended center care (NICHD Early Child Care Research Network, 1999). The coverage of center care and family day care in itself should, given previous findings, exacerbate the association between hours in care and externalizing problems. In a reanalysis of the SECCYD, van IJzendoorn et al. (2004, cited in Belsky et al., 2007) found that time in nonrelative care, and especially center care, was associated with externalizing problems. However, the common use of center care and family day care must be considered in conjunction with the Norwegian regulation of standards for caregiver education and child–caregiver ratios. At least with regard to structural features of child-care quality, Norwegian child-care centers are of relatively high and homogenous quality (Winsvold & Guldbrandsen, 2009), with a maximum adult:child ratio of 3:10 for 1- and 2-year-olds, and 3:19 for those older. In comparison, standards for child-care centers vary considerably across states in the United States, with adult:child ratio standards for 3-year-olds ranging, for instance, from 1:7 to as high as 1:17, with the additional complication that few centers meet these standards (NICHD Early Child Care Research Network, 1999). Given this variability, it is important to note that findings on the influence of child-care quality on associations between child-care quantity and externalizing problems have been mixed, even within the SECCYD (McCartney et al., 2010; NICHD Early Child Care Research Network, 2003). However, with higher structural quality standards in Norway than in the United States, homogenously high-quality care in Norway may help explain the null findings in this study.

It is also of note that Norwegian parental leave policy ensures that most children enter nonparental care near 1 year of age. Findings by the NICHD Early Child Care Research Network (2003) suggest that it is the initiation of high quantities very early in life that have consequences for development. However, in our review of the literature, we could not identify a systematic pattern of associations between child-care quantity and externalizing problems being more common in sociopolitical contexts where children enter nonparental care in their 1st year. In this study, a small proportion of children entered nonparental care in their 1st year, but this did not moderate associations between quantity and problem levels.

In sum, the sociopolitical context, including federal policies of comprehensive early childhood and generous parental leave, is likely one reason hours in child care was unrelated to externalizing problems in this population-based sample of Norwegian children. Near universal access to quality care combined with an age of entry at about 1 year of age may provide a context in which there are, on average, no detrimental consequences of high quantity of care. This conclusion is supported by the work of Love et al. (2003), who draw on evidence from the SECCYD, the evaluation of Head Start, and data from Israel and Australia when they argue that when "standards for good quality care are enforced through government regulator mechanisms, the risk for behavior problems may be explained by factors other than time in care" (p. 1031). Although it is worth noting that a study from Denmark, where early childhood policies are very similar to Norway, demonstrated associations between hours in care at age 3 and behavior problems at age 7 (Gupta & Simonsen, 2010), these authors relied exclusively on models that are statistically most similar to those in which we (and others) also find child-care effects, OLS models in which selection is controlled via child and family covariates and attrition is addressed using listwise deletion.

Study Limitations

A number of limitations to our study must be noted. The baseline participation rate of approximately 40% is a serious issue, for example. This may have introduced bias, with regard to who participated in the study. It is, nonetheless, important to note that these rates are comparable to other population-based studies (Szklo, 1998), and that the baseline participation is 40% of the Norwegian population, excluding mothers giving birth in very small birth units. As a comparison, the NICHD SECCYD had approximately a 50% baseline participation rate (e.g., McCartney et al., 2010), after families within the 10 catchment areas were preemptively excluded from the sampling for a variety of observable risk factors (i.e., characteristics of children, mothers, and communities). Although low participation rates are a serious limitation, it does not seem that this is a greater limitation to our findings than previous findings on this topic. A related limitation is attrition, with missing data for key study variables nearing 50% by 36 months. Despite our efforts to address this issue, including following best-practice recommendations that have yet to become commonplace in this area of study, there remains a risk that our findings were influenced by missing data.

We also note four additional limitations to the study. First, we relied solely on maternal reports of externalizing problems. Although several other studies that have reported associations between maternal reports of child-care quantity and socioemotional outcomes (e.g., Belsky, 1999; Yamauchi & Leigh, 2011), including a study in a Nordic context (Gupta & Simonsen, 2010), evidence of child-care quantity associations with externalizing problems have been less consistent in SECCYD when maternal reports have been compared with caregiver or teacher reports (e.g., NICHD Early Child Care Research Network, 2003). A common trade-off in epidemiological cohort studies is sacrificing multiple informant sources in favor of a very large population-based sample. Nonetheless, given the low correlations observed for maternal- and caregiver-rated behavior problems, our findings may have been different had we been able to examine caregiver reports.

Second, we were not able to examine any later emerging consequences of child-care quantity beyond early childhood. In the SECCYD, 36 months of age was one time point at which there were no associations between child-care quantity and externalizing problems. Confirming our null results at later ages will be useful.

Third, our measures of child-care quantity were restricted to maternal reports at 18 and 36 months. This is problematic if hours reported at these time points do not provide an unbiased representation of amount of time spent in child care during early childhood. This is of particular concern for our estimates of "cumulative" hours, an average of hours spent in care measured at these two time points which is fewer observations of care than used to define cumulative care in the NICHD SECCYD, for example.

Fourth, we did not have assessments of child-care quality, a potential moderator of the effects of hours in care. Given uniform standards for child care in Norway, however, we do not see this as a major weakness of the study. Given such a large sample of care arrangements from a country with homogeneously high-quality care, a null finding averaged across settings is useful. The lack of direct observations of quality data would have been more concerning if, in fact, we had found evidence that quantity, on average, predicted more problems.

Conclusions

Child-care policy in Norway is a significant component of more general progressive family policy designed to promote maternal workforce participation and employment rights as well as universal access to high-quality environments for learning and development beginning in the 2nd year of life (Mandel & Semyonov, 2005; Ministry of Education, 2011). In this sociopolitical context, we found very little support for the hypothesis that a high quantity of care leads to externalizing behavior problems in early childhood. Given public attention to past claims that early child care may pose a developmental risk, our findings should contribute to a reconsideration of existing evidence on this topic with special attention to the methods employed and the sociopolitical context of studies.

References

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2–3 and 1992 Profile*. Burlington: University of Vermont Department of Psychiatry.

Anme, T., & Segal, U. A. (2004). Implications for the development of children in over 11 hours of centre-based care. *Child Care Health and Development, 30*, 345–352. doi:10.1111/j.1365-2214.2004.00429.x

Barnes, J., Leach, P., Malmberg, L.-E., Stein, A., & Sylva, K. (2009). Experiences of childcare in England and socio/emotional development at 36 months. *Early Child Development and Care, 180*, 1215–1229. doi:10.1080/03004430902943959

Bates, J. E., Marvinney, D., Kelly, T., Dodge, K. A., Bennett, D. S., & Pettit, G. S. (1994). Child-care history and kindergarten adjustment. *Developmental Psychology, 30*, 690–700. doi:10.1037/0012-1649.30.5.690

Belsky, J. (1986). Infant day care: A cause for concern? *Zero to Three, 6*, 1–9.

Belsky, J. (1999). Quantity of nonmaternal care and boys' problem behavior/adjustment at ages 3 and 5: Exploring the mediating role of parenting. *Psychiatry-Interpersonal and Biological Processes, 62*, 1–20.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T. (2007). Are there long-term effects of early child care? *Child Development, 78*, 681–701. doi:10.1111/j.1467-8624.2007.01021.x

Borge, A. I. H., & Melhuish, E. C. (1995). A longitudinal-study of childhood behavior problems, maternal employment, and day-care in a rural Norwegian community. *International Journal of Behavioral Development, 18*, 23–42. doi:10.1177/016502549501800102

Brenna, L. R., Bjerkestrand, M., Broström, S., Fagerli, B., Hernes, I., Hornslien, Ø., et al. (2010). *Med forskertrang og lekelyst. Systematisk pedagogisk tilbud til alle førskolebarn* [With urge for research and playfulness. Systematical educational programs for all preschool children] (Research report 2010:8). Oslo, Norway: Ministry of Education.

Campis, L. K., Lyman, R. D., & Prenticedunn, S. (1986). The parental locus of control scale—Development and validation. *Journal of Clinical Child Psychology, 15*, 260–267. doi:10.1207/s15374424jccp1503_10

Carter, A. S., Briggs-Gowan, M. J., Jones, S. M., & Little, T. D. (2003). The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology, 31*, 495–514. doi:10.1023/A:1025449031360

Clarke-Stewart, K. A., & Fein, G. G. (1983). Early childhood programs. In P. H. Mussen, M. Haith, & J. Campos (Eds.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 917–1000). New York: Wiley.

Coté, S. M., Bovin, M., Nagin, D. S., Japel, C., Xu, Q., Zoccolillo, M., et al. (2007). The role of maternal education and nonmaternal care services in the prevention of children's physical aggression problems. *Archives of General Psychiatry, 64*, 1305–1312. doi:10.1001/archpsyc.64.11.1305

Crosby, D. A., Dowsett, C. J., Gennetian, L. A., & Huston, A. C. (2010). A tale of two methods: Comparing regression and instrumental variables estimates of the effects of preschool child care type on the subsequent externalizing behavior of children in low-income families. *Developmental Psychology, 46*, 1030–1048. doi:10.1037/a0020384

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies.

Research in Human Development, 1, 59–80. doi:10.1080/15427609.2004.9683330

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology, 46,* 1454–1480. doi:10.1037/a0020204

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. doi:10.1146/annurev.psych.58.110405.085530

Gupta, N. D., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics, 94,* 30–43. doi:10.1016/j.jpubeco.2009.10.001

Hesbacher, P. T., Rickels, K., Morris, R. J., Newman, H., & Rosenfeld, H. (1980). Psychiatric illness in family practice. *Journal of Clinical Psychiatry, 41,* 6–10.

Hinshaw, S. P. (2002). Process, mechanism, and explanation related to externalizing behavior in developmental psychopathology. *Journal of Abnormal Child Psychology, 30,* 431–446. doi:10.1023/A:1019808712868

Jacob, J. I. (2009). The socio-emotional effects of non-maternal childcare on children in the USA: A critical review of recent studies. *Early Child Development and Care, 179,* 559–570. doi:10.1080/03004430701292988

Jaffee, S. R., Van Hulle, C., & Rodgers, J. L. (2011). Effects of nonmaternal care in the first 3 years on children's academic skills and behavioral functioning in childhood and early adolescence: A sibling comparison study. *Child Development, 82,* 1076–1091. doi:10.1111/j.1467-8624.2011.01611.x

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review, 26,* 52–66. doi:10.1016/j.econedurev.2005.11.005

Love, J. M., Harrison, L., Sagi-Schwartz, A., van IJzendoorn, M. H., Ross, C., Ungerer, J. A., . . . Chazan-Cohen, R. (2003). Child care quality matters: How conclusions may vary with context. *Child Development, 74,* 1021–1033. doi:10.1111/1467-8624.00584

Magnus, P., Irgens, L. M., Haug, K., Nystad, W., Skjaerven, R., & Stoltenberg, C. (2006). Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology, 35,* 1146–1150. doi:10.1093/ije/dyl170

Mandel, H., & Semyonov, M. (2005). Family policies, wage structures and the gender gaps: Sources of earnings inequality in 20 countries. *American Sociological Review, 70,* 949–967. doi:10.1177/000312240507000604

McCartney, K., Bub, K. L., & Burchinal, M. (2006). Selection, reflection, detection. *Monographs of the Society for Research in Child Development, 71,* 105–126. doi:10.1111/j.1540-5834.2006.07103001.x

McCartney, K., Burchinal, M., Clarke-Stewart, A., Bub, K. L., Owen, M. T., & Belsky, J. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology, 46,* 1–17. doi:10.1037/a0017886

Melhuish, E. (2010). Why children, parents and home learning are important. In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford, & B. Taggart (Eds.), *Early childhood matters: Evidence from the effective pre-school and primary education project* (pp. 44–69). London: Routledge.

Ministry of Education. (2006). *Direktiv for familiebarnehager* [Directive for family day care]. Retrieved April 4, 2011, from http://www.lovdata.no/cgi-wift/ldles?doc=/sf/sf/sf-20051216-1555.html

Ministry of Education. (2007). *...og ingen sto igjen. Tidlig innsats for livslang læring.* [...no one left behind. Early efforts to promote lifelong learning] (White Paper No. 16 [2006-2007]).

Ministry of Education. (2010). *Barnehageloven* [The Kindergarten Act]. Retrieved April 4, 2011, from http://www.lovdata.no/all/nl-20050617-064.html

Ministry of Education. (2011). *Early childhood education and care policy.* Retrieved April 4, 2011, from http://www.regjeringen.no/en/dep/kd/Selected-topics/kindergarden/early-childhood-education-and-care-polic.html?id=491283

National Center for Educational Statistics. (2011). *Digest of educational statistics, Table 53.* Retrieved January 9, 2012, from http://nces.ed.gov/programs/digest/d10/tables/dt10_053.asp

NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development, 69,* 1145–1170.

NICHD Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health, 89,* 1072–1077. doi:10.2105/AJPH.89.7.1072

NICHD Early Child Care Research Network. (2002). Child-care structure -> process -> outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13,* 199–206. doi:10.1111/1467-9280.00438

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74,* 976–1005. doi:10.1111/1467-8624.00582

Nilsen, R. M., Vollset, S. E., Gjessing, H. K., Skjaerven, R., Melve, K. K., Schreuder, P., . . . Magnus, P. (2009). Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatric and Perinatal Epidemiology,* 23, 597–608. doi:10.1111/j.1365-3016.2009.01062.x

OECD Early Childhood Education and Care Home Page. (2004). *Country information.* Retrieved November 15, 2012, from http://www.oecd.org/document/57/0,3746,en_2649_39263231_46392953_1_1_1_1,00.html

Romano, E., Kohen, D., & Findlay, L. C. (2010). Associations among child care, family, and behavior outcomes in a nation-wide sample of preschool-aged children. *International Journal of Behavioral Development, 34,* 427–440. doi:10.1177/0165025409351657

SPSS Inc. (2009). *PASW statistics* 17.0 command syntax reference. Chicago, IL: Author.

Statistics Norway. (2011). *Kindergarten statistics*. Retrieved April 4, 2011, from http://www.ssb.no/barnehager/

Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L.-E., & the FCCC-team. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study. *British Journal of Developmental Psychology, 29*, 18–45. doi:10.1348/026151010X533229

Szklo, M. (1998). Population-based cohort studies. *Epidemiologic Reviews, 20*, 81–90.

UNICEF Innocenti Research Center. (2008). *Report Card 8: The child care transition*. Florence, Italy: Author.

van Beijsterveldt, T. C. E. M., Hudziak, J. J., & Boomsma, D. I. (2005). Short- and long-term effects of child care on problem behaviors in a Dutch sample of twins. *Twin Research and Human Genetics, 8*, 250–258. doi:10.1375/1832427054253103

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research,. 16*, 219–242. doi:10.1177/0962280206074463

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of Early Child Care and Youth Development. *Child Development, 81*, 737–756. doi:10.1111/j.1467-8624.2010.01431.x

Vandell, D. L., & Corasaniti, H. (1990). Child care and the family: Complex contributors to child development. *New Directions for Child Development, 49*, 23–37. doi:10.1002/cd.23219904904

Winsvold, A., & Guldbrandsen, L. (2009). *Kvalitet og kvantitet—kvalitet i en barnehagesektor i sterk vekst* [Quality and quantity—Quality in a child-care sector in rapid growth] (Research Report 2/2009). Oslo, Norway: NOVA.

Yamauchi, C., & Leigh, A. (2011). Which children benefit from non-parental care? *Economics of Education Review, 30*, 1468–1490. doi:10.1016/j.econedurev.2011.07.012

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Robustness Checks.

Table S1. Correlations Between Measures of Externalizing Problems and Hours in Child Care at 18 and 36 Months.

Figure S1. Unconditional Associations Between Hours in Child Care and Externalizing Problems at 18 and 36 Months of Age.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1171–1190

Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics

Tran D. Keys, George Farkas,
Margaret R. Burchinal, Greg J. Duncan,
Deborah L. Vandell, Weilin Li,
and Erik A. Ruzek
University of California, Irvine

Carollee Howes
University of California, Los Angeles

This article examines associations between observed quality in preschool center classrooms for approximately 6,250 three- to five-year-olds and their school readiness skills at kindergarten entry. Secondary analyses were conducted using data from four large-scale studies to estimate the effects of preschool center quality and interactions between quality and demographic characteristics and child entry skills and behaviors. Findings were summarized across studies using meta-analytic methods. Results indicate small, but statistically significant associations for preschool center quality main effects on language and mathematics outcomes with little evidence of moderation by demographic characteristics or child entry skills and behaviors. Preschool center quality was not reliably related to socioemotional outcomes. The authors discuss possible explanations for the small effect sizes and lack of differential effects.

This article examines associations between the child-care center quality experienced by preschool-aged children (3–5 years old) and the school readiness skills of these children at kindergarten entry, and asks whether those associations are moderated by either demographic characteristics or the child's entry skills and behaviors. School readiness encompasses physical, cognitive, language, and behavioral aspects of development (Kagan, Moore, & Bredekamp, 1995). This study examines the following school readiness outcomes: language, mathematics, social skills, and externalizing problem behaviors using data from four large multisite child-care studies.

Children typically spend many hours in child-care settings during early childhood (Pianta, Burchinal, Barnett, & Thornburg, 2009) and the quality of those settings is thought to be an important factor in young children's development (Lamb & Ahnert, 2006). High-quality settings for 3- to 5-year-olds include frequent warm and stimulating interactions between caregivers and children and clear intentional instruction (Pianta et al., 2009). Several child-care quality assessment tools have been developed and are widely accepted. As measured by these instruments, child-care quality is often positively associated with child outcomes, but estimated effects range from very small to moderate depending on the study (Burchinal, Kainz, & Cai, 2011; Camilli, Vargas, Ryan, & Barnett, 2010; Pianta et al., 2009; Vandell, 2007).

Whether child-care quality is related to developmental outcomes differently for children with risk factors is an important policy question. Both federal and state governments invest heavily in providing (presumably high-quality) child care as a means to improve school readiness skills for children from economically disadvantaged families. However, there is mixed support in the literature for the notion that high-quality child care is more beneficial for particular subgroups of children.

The goals of this article are to examine the magnitude of preschool center quality main effects and the existence of moderators of these effects by providing integrated analyses across four large-scale

---

The authors thank Anamarie Auger, Thurston Domina, Aleksandra Holod, Jerry Keys, and four anonymous reviewers for their helpful comments and suggestions on previous versions of this manuscript. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090467 to the University of California, Irvine (PI: George Farkas). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Tran Keys, Department of Education, University of California, Irvine, 3200 Education, Irvine, CA 92697-5500. Electronic mail may be sent to tran.keys@uci.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0006
DOI: 10.1111/cdev.12048

child-care studies. Our two research questions are as follows: (a) to what extent are children's school readiness skills higher when they experience higher quality preschools, and (b) do the effects of preschool center quality on school readiness skills differ by child demographic background or child entry characteristics such as cognitive skills, attention skills, and problem behaviors? Our examination of child entry skill levels and behaviors as potential moderators of child-care quality effects may be of particular interest to the field due to the lack of prior research in this area.

Developmental Perspectives on Child-Care Quality

Our understanding of early childhood development is rooted in several complementary theoretical perspectives. A central proposition of bioecological (Bronfenbrenner & Morris, 2006) and life course theory (Elder, 1996) is that experiences provide the basis for development and early experiences can lay the foundation for later development. Prior studies highlight the importance of early childhood for brain development (Knudsen, Heckman, Cameron, & Shonkoff, 2006; Sapolsky, 2004) and subsequent academic and social development (Shonkoff & Phillips, 2000). Economic models of human capital development suggest the cumulative role of cognitive and socioemotional skills, as well as skill investments made by families, preschool programs, and schools in producing human capital (e.g., Cunha, Heckman, Lochner, & Masterov, 2006).

The literature on child-care quality and outcomes provides empirical evidence for these theories. The experimental literature suggests that higher quality care can have a lasting impact on development when that care begins during the early childhood years (Belfield, Nores, Barnett, & Schweinhart, 2005; Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001). The observational literature also suggests modest but relatively consistent evidence that higher quality child care is related to improved cognitive and social development (Burchinal et al., 2011; Peisner-Feinberg et al., 2001; Vandell, 2007; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010). There is some, albeit mixed, evidence of "active range" thresholds in which gains in outcomes are larger at higher levels of quality (Burchinal, Vandergrift, Pianta, & Mashburn, 2009; Vandell et al., 2010). In recent analyses that combined results across five early child-care studies (Burchinal et al., 2011), children in lower quality care had lower outcomes than children in higher quality care, but the nonlinear association between quality and outcomes was only observed in higher quality care and with mostly instructional (rather than emotional) support quality measures.

Magnitude of Child-Care Quality Main Effects

The magnitude of associations between child-care quality and school readiness skills vary, depending on the study and the domain of development examined. Prior large-scale observational studies that did not involve random assignment to early child-care conditions (i.e., quasi-experiments) and that included larger and more representative samples than experimental studies, provide some evidence of associations between higher quality child care and positive, particularly cognitive, child outcomes (Burchinal & Cryer, 2003; Burchinal et al., 2011; Gormley, Gayer, Phillips, & Dawson, 2005; Howes et al., 2008; Mashburn et al., 2008; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2000, 2002, 2006; NICHD Early Child Care Research Network & Duncan, 2003; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Reynolds, Temple, Robertson, & Mann, 2002; Vandell et al., 2010; Votruba-Drzal, Coley, & Chase-Lansdale, 2004).The effect sizes in these studies vary, with larger effects observed in studies with the greatest range in the quality of settings observed, as in the cost, quality, and child outcomes study (Peisner-Feinberg & Burchinal, 1997) and the NICHD SECCYD (NICHD Early Child Care Research Network, 2003, 2006).

Research on the effects of nonmaternal child care on children's socioemotional development has produced mixed results. On one hand, some research suggests that for some children, care outside the home can be disadvantageous to their social development (Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; NICHD Early Child Care Research Network, 2002, 2003). On the other hand, some research suggests that child-care programs can benefit children's socioemotional development, especially when quality is high and among children from low-income families (Votruba-Drzal et al., 2004). There is research that suggests that nonmaternal child care may benefit low-income children's social development by compensating for the negative behavioral effects children receive from being raised in a low-quality home environment (Watamura, Phillips, Morrissey, McCartney, & Bub, 2011).

Of particular interest are prior findings regarding the links between child-care quality and school readiness outcomes that have resulted from analyses of the four studies analyzed in the current study: the NICHD Study of Early Child Care and Youth Development (SECCYD), the Early Childhood Longitudinal Study–Birth Cohort (ECLS–B), the National Center for Early Development and Learning (NCEDL) 11-state prekindergarten evaluation, and the Early Head Start (EHS) study. Of the four, the NICHD SECCYD has been most often analyzed. The prevailing finding from these analyses is that the quality of care is consistently but modestly related to improved cognitive and language outcomes at 15, 24, 36, and 54 months, even after controlling for multiple child and family characteristics (NICHD Early Child Care Research Network, 2000, 2002, 2006; NICHD Early Child Care Research Network & Duncan, 2003). The NCEDL 11-state evaluation of prekindergarten programs also reported modest associations between quality and various child outcomes (Burchinal et al., 2008; Burchinal et al., 2009; Howes et al., 2008; Mashburn et al., 2008). Research using the ECLS-B and EHS studies has shown positive impacts on child outcomes of both non-maternal early childhood education program participation (Bassok, 2010) and higher quality child-care centers (Love et al., 2003; Love et al., 2005).

A recent meta-analysis of the published literature examining the association between child-care quality and child cognitive, academic, and socioemotional outcomes yielded a modest, albeit statistically significant, partial correlation of $r = .12$ (Burchinal et al., 2011). Burchinal et al. (2011) found stronger associations for language and academic outcomes than for socioemotional outcomes, and for outcome measures that were more closely aligned with the measures of child-care quality. Their reanalyses focused on the association between child-care quality and child outcomes among low-income children only, whereas this study includes all children in center-based care during the preschool period. Also, their study did not examine moderators. Along with examining demographic moderators, a contribution of our study to the child-care quality field is inclusion of *child entry skills and behaviors* as possible moderators of associations between observed child-care quality and child outcomes.

Existence of Child-Care Quality Moderators

Prior research yields little consensus on the existence of variables that moderate the effects of child-care quality on child outcomes. When moderators have been examined, the focus has been on demographic moderators such as race or ethnicity, gender, and socioeconomic status. Most studies of the NICHD SECCYD report main effects, but find few moderating effects (Vandell, 2007). Similarly, analyses of three large child-care studies not only found that child-care quality was related to development for all children but also found at best limited evidence that this effect was larger for at-risk children (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000).

Researchers have postulated competing hypotheses regarding differential effects of child-care programs on child outcomes. Especially relevant to children's participation in high-quality early education programs is the *compensatory hypothesis* (Sameroff & Chandler, 1975), which predicts that children who are at risk because of their particular race or ethnicity or gender, or who are at risk because of economic disadvantage, low skills, or difficult temperaments are likely to derive greater benefits from high-quality early education programs relative to children who are not at risk. This hypothesis provides the rationale for the continued funding of programs such as Head Start. Alternatively, the *accumulated advantages* and *skill begets skill hypotheses* (also sometimes referred to as the Matthew effect—*the rich get richer*) posits that children who are more advantaged due to family or neighborhood circumstances or who have greater initial individual abilities will derive greater benefits from high-quality early education programs than their less advantaged peers because of their ability to build on these advantage or skills (Cunha et al., 2006). To examine these competing hypotheses, we test interactions between preschool center quality and three demographic characteristics (race or ethnicity, gender, maternal education) as well as between preschool center quality and three child entry characteristics (cognitive skills, attention skills, and externalizing problem behaviors).

Demographic Characteristics

We examined three demographic moderators: race or ethnicity, gender, and socioeconomic status (SES). Regarding race or ethnicity as a moderator, in line with the compensatory (vs. accumulated advantages) hypothesis, stronger child-care effects have been reported for African American and Hispanic children than for White children in an analysis of nationally representative data (Magnuson, Ruhm, & Waldfogel, 2007), an evaluation of a prekindergarten program (Gormley et al., 2005), and a study relating quality to outcomes across three early child-care studies (Burchinal et al., 2000). There is mixed evidence that the effects of early childhood programs are moderated by child gender. Two studies reported that boys appear to benefit more than girls from child-care experiences (Brooks-Gunn, Han, & Waldfogel, 2002; Votruba-Drzal et al., 2004), while reanalysis of the Abecedarian and Perry Preschool Projects found the opposite (Anderson, 2008). If the effect of high-quality child care is compensatory, boys will profit more than girls because boys tend to be less developmentally advanced than girls, putting them at greater risk for poor educational outcomes (Matthews, Ponitz, & Morrison, 2009). Finally, prior studies have found some support for moderation by maternal education and other measures of SES such as income level or poverty status. The cost, quality, and outcomes study found larger child-care quality effects for children whose mothers had low educational levels (Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001), and analyses of the NICHD SECCYD suggested that economically disadvantaged children seem to benefit more from better child-care quality (Dearing, McCartney, & Taylor, 2009).

Child Entry Skills and Behaviors

As mentioned earlier, the literature has posited a compensatory hypothesis (Sameroff & Chandler, 1975), in which children with low skills or difficult temperaments are likely to benefit the most from high-quality early education programs. But the opposite hypothesis is generated by the skill begets skill hypothesis, where the productivity of child-care investment is greatest for children with higher initial abilities (Cunha et al., 2006). We were not able to find prior empirical work that examined differential effects of child-care quality on children of varying *entry* skills and behaviors. In this study we test three child entry skills and behaviors as possible moderators of preschool center quality effects: cognitive skills, attention skills, and externalizing problem behaviors. For each of these, we test whether children who are low on the variable experience stronger effects of high-quality child care (compensatory effect) or whether children who are high on the variable experience stronger effects of high-quality child care (skill begets skill).

This Study

This study extends prior work in several ways. First, perhaps for the first time, we examine several child entry skills and behaviors as possible moderators of associations between observed child-care quality and child outcomes. Second, we formally estimate quality main effects and interactions across several of the largest child-care studies. As in a few prior childhood education studies (Burchinal et al., 2011; Duncan et al., 2007; Early et al., 2007), parallel analyses are conducted across multiple studies to ensure that the findings from each study do not differ simply because of modeling differences. Most importantly, we include a final statistical test that formally combines the findings across studies to test our hypotheses. Thus, as in Duncan et al. (2007), we apply meta-analytic techniques to summarize the results of our own secondary analyses of four large-scale studies, each of which had comparable regression models applied to it.

Research Hypotheses

Consistent with the main effects literature outlined earlier, we predicted small positive preschool center quality main effects for all school readiness outcomes. With regard to differential effects, we predicted that in accordance with the compensatory hypothesis, children will benefit more from high-quality preschool programs if they are African American or Hispanic, male children, if their mothers had low levels of education, or if they enter preschool with moderately low cognitive skills, low attention skills, or more externalizing behavior problems. We also examine the accumulated advantages and skill begets skill hypotheses that the opposite is true—that the effects of high-quality preschool programs are weakest for high-risk children and strongest for those children who begin the program with the most productive skills and behaviors.

Method

The four databases analyzed in the study were selected because they are large child-care studies containing the necessary preschool center quality and child outcome measures. Table 1 provides descriptions of these databases.

Sample and Procedure

NICHD Study of Early Child Care and Youth Development

Nonexperimental longitudinal data from the NICHD SECCYD are drawn from a multisite study

Table 1
Descriptions of the Four Studies

	NICHD SECCYD	ECLS-B[a]	NCEDL 11-state	EHS
Sample				
All children in center care	733	1,400[b]	2,995	1,140
Children with observed child-care quality scores	670	543	2,982	676
Number of classrooms	623	543	721	n/a
Year quality & posttest collected	1995–1996	2005–2006	2001 for Multistate study of pre-K and 2004 for SWEEP[c]	2001–2003 (three cohorts)
Population analysis sample represents	Children at the 10 locations across the United States who were in center-based care the year before pre-K	Children born in 2001 who were in center care at 4 years of age for at least 4 months and 10 hr/week	State funded pre-K classrooms and children in 11 participating states	Children who had been in EHS evaluation study as infants or toddlers (includes both treatment & control)
Percent Head Start classroom	9	24	15	45
Mean child (*SD*) age (mos) at pretest assessment	37.68 (0.75)	24.59 (1.41)	55.56[c] (3.84)	37.12 (1.42)
Mean child (*SD*) age (mos) at outcome assessment	56.86 (1.11)	54.37 (4.33)	60.60[c] (3.84)	59.00 (3.64)

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start.

[a] All *N*s for the ECLS–B are rounded to the nearest 50 per NCES/IES reporting requirements. [b] A random subset of the ECLS–B sample had child-care settings evaluated, and our analyses subsample includes children who meet our ≥ 4 month criteria and were in care for at least 10 hr/month. [c] SWEEP = State-Wide Early Education Programs. As with many studies of child-care programs, the NCEDL project was not able to gain access to classrooms during the first month of the year so "fall" test scores were collected in the second and occasionally the third months of the program. The project needed to collect the "spring" assessments prior to the last month because so many special events and graduation events are scheduled for that time.

of births in 1991 (visit http://secc.rti.org/ for more information on NICHD SECCYD or see NICHD Early Child Care Research Network, 2006). Participants were recruited from hospitals located at 10 sites across the United States. Although not nationally representative, the study sample (n = 733, children in center-based child care only) closely matches national and census tract records with respect to demographic variables such as ethnicity and household income. Family, child care, and child characteristics were measured in home- and lab-based assessments when the children were 6, 15, 24, 36, and 54 months. Through observations, interviews, questionnaires, and child assessments, data include repeated assessments of demographic and parental characteristics; quality of parenting; type, amount, and quality of child care; and children's cognitive, academic, and social skills. The study's rich information collected on the child's demographic, home, and care environment allow researchers to control for a host of confounds related to children selecting into different child-care settings.

Early Childhood Longitudinal Study–Birth Cohort

The ECLS–B followed a large, nationally representative sample of children born in 2001 to kindergarten entry (for more information on ECLS–B, visit http://nces.ed.gov/ecls/birth.asp). This study provides detailed information on children's development, health, and learning experiences during the years leading up to school entry. Data were collected from sample children and their parents, as well as child-care providers and teachers. Children's behavioral, cognitive, social, and physical abilities were directly assessed at 9, 24, and 48 months (preschool), and at kindergarten entry. The base year sample includes about 10,700 infants with data collected from birth certificate records and parent-provided information during the 9-month data collection wave. Data collection continued at 24 and 48 months and ended during the school year in which the child attended kindergarten, with an approximate completed sample size of 8,000. The primary child-care setting was observed for a randomly selected subset of the sample at 24 and 48 months. To address the concern that children in our ECLS–B sample may not have had adequate exposure to preschool to consider its influence, we restrict our ECLS–B analysis sample to children who were in 4 or more months in child-care settings for at least 10 hr/month (n = 1,400 approximately).

National Center for Early Development and Learning

The NCEDL followed nearly 3,000 (n = 2995) children enrolled in 721 pre-K classrooms randomly chosen within selected regions of 11 states with mature pre-K programs (for more information on NCEDL, visit http://fpg.unc.edu/projects/national-center-early-development-learning or see Early et al., 2005). These 11 states served approximately 80% of children in the United States, who attended state pre-K programs in the study years of 2001–2003. The design involved random sampling of 40 pre-K sites in six states during the 2001–2002 school year and 100 sites within an additional five states during 2003–2004. In both periods, within each pre-K site, one classroom was randomly selected to participate, and four children were randomly selected per classroom among children whose parents signed and returned the consent form, were scheduled to enter kindergarten the next year, spoke English or Spanish at home, and did not have an Individualized Education Plan. Children and their classroom experiences were assessed in the fall and spring of the pre-K year for all children and in the fall and spring of the kindergarten year for the first cohort of children.

Early Head Start

Our final study comes from the evaluation of the EHS program (visit http://www.acf.hhs.gov/programs/opre/research/topic/overview/early-head-start for more information). Although the EHS study is based on random assignment into early care, this study uses both treatment and control data on the child care that parents selected for their children between the ages 36 and 60 months (n = 1,140). The control group was a "services as usual group" and could attend whatever was available in the community, including child care that was not provided by the EHS program. In 1996, approximately 3,000 children under 1 year of age from low-income families from 17 sites were randomly assigned to receive EHS services, or to a control group. Treatment ended when the child was 3 years old and children in both groups were followed into kindergarten. Data on children, families, and child-care arrangements were collected when the children were 14, 24, and 36 months old, and in the spring just prior to kindergarten entry in the form of direct assessments of the children, maternal report interviews and questionnaires, and observations of the quality of the home and child-care environments.

Measures

Table 2 describes the key dependent and independent variables in our analyses along with information on when they were measured and their Cronbach's alpha (where available). Table 3 lists descriptive statistics of the key variables of interest in each of the studies. The measures of interest for each of the four studies are described next.

Child-Care Quality Measures

We include in the study the three child-care quality measures described next that assess practices thought to improve children's cognitive, achievement, and socioemotional skills; these include two global quality measures and a measure of instructional quality.

The Early Childhood Environment Rating Scale–Revised (ECERS–R; Harms, Clifford, & Cryer, 1998)—used in the ECLS–B, NCEDL, and EHS studies—is a widely used measure of global classroom quality, specifically designed for use in classrooms serving children between 2½ and 5 years of age. The ECERS–R has been used in numerous studies over the past 25 years and has been shown to correlate with child outcomes (e.g., Peisner-Feinberg et al., 2001). The ECERS–R includes 37 items with six subscales measuring the following: quality of space and furnishings, personal care routines, language and reasoning opportunities, activities, interactions, and program structure. Once in the fall

Table 2
Key Dependent and Independent Variables

	NICHD SECCYD	ECLS–B	NCEDL 11-state	EHS
Outcomes				
Language	Preschool Language Scale–3 (54 mos; α = .89–.92)	ECLS–B literacy (composite of PPVT & other measures; 48 mos; α = .91)	PPVT–III (spring pre-K; α = .92–.98)	PPVT–III (63 mos; α = .92–.98)
Mathematics	WJ–Revised applied problems (54 mos; α = .91)	ECLS–B math (math ECLS-K developed measures; 48 mos; α = .89)	WJ III applied problems (spring pre-K; α = .92–.94)	WJ III applied problems (63 mos; α = .92–.94)
Social skills	California Preschool Social Competency Scale (caregiver rating; 54 mos; α = .86)	ECLS–B social skills (composite created from several items of caregiver rating; 48 mos; 7 items, α = .98)	Teacher–Child Rating Scale—Social Skills (spring pre-K; α = .95)	n/a
Externalizing problem behaviors	CBCL–externalizing problem behaviors only (caregiver rating; 54 mos; α = n/a)	ECLS–B externalizing problem behaviors (composite created from several items; 48 mos; 10 items, α = .98)	Teacher–Child Rating Scale conduct problems subscale (spring pre-K; α = .92)	CBCL aggressive behaviors only (63 mos; α = n/a)
Preschool center quality	ORCE (36, 54 mos; α = .80–.90)	ECERS–R composite (48 mos; α = .92)	ECERS–R composite (α = .90) & CLASS emotional climate (α = .85–.86) & CLASS instructional climate (α = .85–.87; pre-K)	ECERS–R composite (48 mos; α = .92)
Child entry skills & behaviors				
Cognitive skills	Bayley (24 months; α = .83)	BSID–SF (a modified version of the Bayley; 24 mos; α = .80)	PPVT–III (Fall pre-K; α = .92–.98)	Bayley (36 mos; α = .83)
Attention skills	CBCL attention subscale (24 mos; α = n/a)	Bayley Infant Behavior Record (IBR) attention rating (24 mos; 4 items, α = .93)	Teacher–Child Rating Scale task orientation subscale (fall pre-K; α = .95)	Three-bag assessment of sustained attention to objects (36 mos; α = n/a)
Externalizing problem behaviors	CBCL–externalizing problem behaviors only (36 mos; α = .86)	n/a	Teacher–Child Rating Scale conduct problems subscale (fall pre-K; α = .91)	n/a

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start; PPVT–III = Peabody Picture Vocabulary Test, 3rd edition; WJ = Woodcock–Johnson; CBCL = Child Behavior Checklist; ORCE = Observational Record of the Caregiving Environment; ECERS–R = Early Childhood Environment Rating Scale–Revised; CLASS = Classroom Assessment Scoring System; All reliability coefficients (alphas) are from the measure's authors except for the ECLS–B, in which the externalizing problem behaviors and social skills scales were created using confirmatory factor analysis.

and once in the spring, a trained observer spends several hours observing the classroom on a typical day, noting all aspects of the environment including materials, safety, health, language interactions, discipline, and relationships. Scores on the ECERS–R range from 1 to 7 with 1 = *inadequate*, 3 = *minimal*, 5 = *good*, and 7 = *excellent* quality.

The Observational Record of the Caregiving Environment (ORCE; NICHD Early Child Care Research Network, 2000) was designed specifically

Table 3
Descriptive Statistics

	NICHD SECCYD			ECLS–B[a]			NCEDL 11-state			EHS		
	(Only center care)			(Only center care for ≥4 months, & 10 + hr/month)			(Study is all center care)			(Only center care)		
	N	*M*	*SD*	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Outcomes												
Preschool language scale	690	101.78	17.84									
Receptive vocabulary measure (i.e., PPVT)				1,300	8.85[c]	1.92	2,435	96.30	14.31	956	92.44	15.29
WJ applied problems	690	104.61	15.21				2,273	99.11	12.85	1,010	89.07	20.09
Math ECLS–K developed measures				1,250	24.29	7.00						
CBCL–externalizing problem behaviors	592	49.41	9.61							1,120	11.03	6.78
ECLS–B externalizing problem behaviors				1,350	.16	.96						
Hightower–conduct problems subscale							2,634	1.57	0.75			
CA preschool social competency scale	583	92.54	12.37									
ECLS–B social skills				1,250	−.06	1.02						
Hightower–social skills scale							2,644	3.64	0.77			
Child entry skills & behaviors												
Cognitive skills	708	93.60	14.38	1,250	127.11	11.10	2,298	94.00	15.02	1,037	91.22	12.62
Attention skills	710	2.99	0.66	1,250	13.62	3.74	2,566	3.52	0.94	1,053	4.89	0.97
Externalizing problem behaviors	402	26.65	18.83				2,571	1.51	0.53			
Observed process quality												
ORCE (36 months)	244	2.74	0.44									
ORCE (54 months)	670	3.04	0.55									
ECERS–R composite				550	4.51	1.10	2,967	3.84	0.81	676	5.25	1.14
CLASS instructional climate							2,927	2.07	0.84			
CLASS emotional climate							2,927	5.56	0.68			
Demographics												
Ethnicity total *N*	733			1,300			2,898			1,140		
White		0.80			0.52			0.41			0.37	
Black		0.11			0.18			0.18			0.33	
Hispanic		0.05			0.23			0.26			0.24	
Asian[b]					0.02							
Other (not Black, Hispanic, White)		0.04			0.05			0.14			0.06	
Gender total *N*	733			1,300			2,966			1,140		
Male		0.50			0.50			0.49			0.51	
Mother's education total *N*	733			1,400			2,885			1,140		
Mom ed: 12 years or less		0.24			0.42			0.59			0.74	
Mom ed: Some college		0.57			0.27			0.23			0.25	
Mom ed: BA plus		0.19			0.31			0.18			0.01	

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study– Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start; PPVT–III = Peabody Picture Vocabulary Test, 3rd edition; WJ = Woodcock–Johnson; CBCL = Child Behavior Checklist; ORCE = Observational Record of the Caregiving Environment; ECERS–R = Early Childhood Environment Rating Scale–Revised; CLASS = Classroom Assessment Scoring System; ECLS–B weight W31C0 was used for all ECLS–B variables except for the ECERS–R composite mean and standard deviation, which was weighted by W33P0. All *N*s for the ECLS–B are rounded to the nearest 50 per NCES/IES reporting requirements.

[a]All *N*s for the ECLS–B are rounded to the nearest 50 per NCES/IES reporting requirements. [b]Asian category only in ECLS–B (the other three studies include Asians in other category). [c]The mean for the ECLS–B receptive vocabulary test is the average number of correct items. From the manual, it "is based on 15 (of 16) PPVT items, which also tested receptive vocabulary (children were asked to select one of four pictures that represented the stimulus word). One PPVT item was removed from scoring due to the presence of differential item functioning."

for the NICHD SECCYD to assess the quality of caregiver–child interaction experienced by individual children. Observations of child-care quality are made during a minimum of four 44-min observation cycles spread over a 2-day period. At the end of each 44-min cycle, observers record ratings from

1 = *not at all characteristic* to 4 = *highly characteristic* to describe caregiver behavior. A positive caregiving composite score is created by obtaining a mean score across scales over all of the ORCE cycles at a given age period. Higher scores indicate that caregivers are more sensitive to children's behaviors, provide greater cognitive stimulation, have warmer and more positive demonstrations, foster greater explorations, and are less emotionally detached. Scores range from 1 (*unresponsive or harsh caregiving*) to 4 (*frequent responsive and stimulating caregiving*).

Finally, the NCEDL also used an early version of the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), an observational assessment to rate teacher–child interactions on nine dimensions of the socioemotional and instructional climate of the classroom. During 2 observation days in each of the fall and spring, observers rated the pre-K classroom and teacher in repeated cycles of roughly 30 min. Observation days lasted from the time children arrived until they started nap in full-day programs or left for the day in half-day programs. Quality ratings of each dimension are on a scale of 1–7 with 1–2 indicating low levels, 3–5 medium levels, and 6–7 high levels. A score for each dimension was created by averaging across all cycles within a data collection period. Factor analysis of the CLASS yielded two factors—instructional climate and emotional climate (Early et al., 2005). Prior analyses of the NCEDL data (Burchinal et al., 2009; Burchinal et al., 2011) indicated that the CLASS Instructional Climate was significantly associated with academic outcomes and CLASS Emotional Climate was significantly associated with emotional outcomes; therefore, we used the CLASS Instructional Climate in the analyses with language and mathematics outcomes and CLASS Emotional Climate in the analyses with social skills and externalizing problem behaviors outcomes.

Outcome Measures

This study examined four outcome measures—language, mathematics, social skills, and externalizing problem behaviors.

Language. Several measures of receptive vocabulary were used across the studies. The Peabody Picture Vocabulary Test, 3rd edition (PPVT–III; Dunn & Dunn, 1997), used in the NCEDL and EHS studies, is an achievement test of receptive vocabulary that relates to other measures of language, literacy, and academic achievement. Children are shown a set of four pictures and asked to select the picture that best represents the meaning of a word spoken by the examiner. The Preschool Language Scale–3 (Zimmerman, Steiner, & Pond, 1992), used in the NICHD SECCYD, measures a range of language behaviors, including vocabulary, morphology, syntax, and integrative thinking, grouped into two subscales: auditory comprehension and expressive language. The ECLS–B research team created a new instrument based on item response theory analyses of items from language tests such the PPVT and literacy tests such as the Woodcock–Johnson (Snow et al., 2009).

Mathematics. The NCEDL and EHS studies administered the Woodcock–Johnson (WJ) III applied problems subtest (Woodcock, McGrew, & Mather, 2001) to measure early mathematics skills. Similar to the PPVT, the WJ subtests are standardized measures that are widely used and endorsed by early childhood practitioners to assess children's progress over time in basic skill areas. The WJ Applied Problems subtest examines the individual child's ability to analyze and solve math problems. The NICHD SECCYD study used the Woodcock–Johnson revised (Woodcock & Johnson, 1989) measure. The ECLS–B research team created a new math achievement test based on item response theory analyses of items from major math achievement tests including the WJ (Snow et al., 2009).

Social skills. For consistency across studies, even though some of the studies contained both maternal and nonmaternal caregiver reports, we included only the nonmaternal caregiver measures as all of the studies contained nonmaternal ratings, but not all contained maternal ratings. The California Preschool Social Competency scale (Levine, Elzey, & Lewis, 1969) was used in the NICHD SECCYD and is a 30-item questionnaire that assesses peer interaction skills as well as the child's ability to follow instructions and communicate effectively. We created a composite from seven items for the ECLS–B social skills outcome measure. The Teacher–Child Rating Scale (TCRS)–Social Skills (Hightower et al., 1986) was examined in the NCEDL. The EHS study did not contain a caregiver rating that was comparable to the other three outcomes measures; therefore, we did not include the social skills outcome for EHS.

Externalizing problem behaviors. Children who exhibit externalizing problem behaviors direct emotional problems outward into aggressive behaviors (Achenbach, 1991). Similar to the social skills outcomes, even though some of the studies contained both maternal and nonmaternal caregiver reports, we included only the nonmaternal caregiver measures as all of the studies contained nonmaternal ratings but not all contained maternal ratings. The Child Behavior Checklist (CBCL; Achenbach, 1991) is a caregiver-report questionnaire on which the child is rated on various competencies and behavioral problems. The CBCL–externalizing problem behaviors subscale outcome measure was used in the NICHD SECCYD and EHS study and includes aggressive behaviors such as arguing, screaming, seeking attention, teasing others, behaving in a threatening manner, and displaying a temper. We created a composite from 10 items for the ECLS–B externalizing problem behaviors outcome measure. The TCRS conduct problems subscale (Hightower et al., 1986; with six items including child is "disruptive in class" and "overly aggressive") was examined in the NCEDL.

Child Entry Skills and Behaviors

This study examined three child entry skills and behaviors—cognitive skills, attention skills, and externalizing problem behaviors.

Cognitive skills. The Bayley Scales of Infant Development (in the NICHD SECCYD, EHS studies) or a modified version of the Bayley (in the ECLS–B study; Bayley, 1993) was administered at 24 months for the NICHD SECCYD and ECLS–B studies and 36 months for EHS to assess general cognitive development. The ECLS–B used a modified version of the Bayley—the BSID–Short Form (SF; Snow et al., 2009). It contained a subset of the items in the full Bayley, but the psychometric properties are similar and scores from the BSID–SF are compatible with the full Bayley. The PPVT–III (Dunn & Dunn, 1981) was administered at the beginning of preschool in the NCEDL study (see the previous section for a description of this measure).

Attention skills. The four studies administered a variety of measures of attention skills. CBCL (Achenbach, 1991) was used in the NICHD SECCYD study and includes an attention subscale measured at 24 months. The CBCL attention subscale assesses attention problems potentially indicative of concentration difficulties, problems sitting still, impulsivity, nervousness, and poor performance in school. The ECLS–B included attention ratings at 24 months from the Bayley Infant Behavior Record (IBR), reflecting the extent to which the infant is responsive to the examiner and to test materials (Snow et al., 2009). The TCRS (Hightower et al., 1986), used in the NCEDL study, is a behavioral rating scale that assesses children's social competence and problem behaviors. The task orientation subscale (with five items including child is "well organized" and "completes work") is used in this study. The Three-Bag Assessment measuring child's sustained attention to objects was measured at 36 months in the EHS.

Externalizing problem behaviors. Of the four studies, only the NICHD SECCYD and NCEDL had well-defined or readily available externalizing problem behavior measures at preschool entry to include in the analysis. For the NICHD SECCYD, the CBCL externalizing behavior subscale at 36 months was included in the analyses. The TCRS conduct problems subscale (Hightower et al., 1986) assessed in the fall of pre-K in the NCEDL is used in this study. Description of the child entry measures can be found in the externalizing problem behavior outcome measure section above.

Control Variables

Extensive covariates were included to minimize selection bias. Where available, our control variables include the following: pretest school readiness measures, child age, low birth weight, maternal characteristics (such as depression and sensitivity), parenting quality, household structure, and family income. Most control variables in the studies reflect demographic characteristics and social risk factors collected during interviews with the mother. Our pretest measures of the variables used as school readiness outcomes should be particularly effective at eliminating biases from selection into child-care quality.

Missing Data

In these studies, as in all longitudinal studies, information is missing on some of the variables. We dealt with this by using multiple imputations with Stata software to fill in missing variables based on regressions of the available variables, plus a randomly selected error term. Each database is then used to produce coefficient estimates, which are then combined into a single estimate with an appropriately estimated standard error. Under the assumption that data are missing at random, which can typically be met using the large number of controls available in these studies, this methodology has been shown to produce estimates with excellent statistical properties (Allison, 2002).

Analytic Approach

For each study, we at least partially accounted for selection bias by including the child's pretest skills level as a covariate, along with a host of other covariates listed in Table 3 and in the online appendix. This amounts to estimating "residualized change models" (NICHD Early Child Care Research Network & Duncan, 2003) in which a particular child school readiness skill at the end of child care is regressed on the pretest skill, average child-care quality, plus a full set of demographic and child controls. Our school readiness outcomes are language, mathematics, social skills, and externalizing problem behaviors at kindergarten entry. Our predictors include observed child-care quality, interactions between child-care quality and demographic characteristics, and interactions between child-care quality and child entry cognitive skills, attention skills, and externalizing problem behaviors.

To be as comprehensive as possible in testing both the compensatory and accumulated advantages or skill begets skill hypotheses, we estimated models interacting child-care quality with the child entry skills and behaviors in models containing both continuous and categorical child entry skills and behaviors variables—Quality × Continuous Entry Skill or Behavior, Quality × Low Entry Dummy, and Quality × High Entry Dummy. Both approaches were used because the cross-product interactions between quality and child characteristics as continuous variables will always provide the most powerful approach to detecting interactions when the underlying assumptions are met (DeCoster, Iselin, & Gallucci, 2009). That is, this approach is always preferable when the interaction reflects a linear increase (or decrease) in the association between quality and child outcomes with an increase in child characteristics. In contrast, this approach will fail to detect interactions when the slope between quality and child outcomes is qualitatively different when child outcomes are initially either high or low. To examine this possibility, we also created categorical variables that reflected whether child characteristics were conceptually high or low in the pretest score. We chose this conceptual approach over a more traditional approach that depended on a quartile split to ensure that what constituted a high level or low level on the child characteristics was clinically meaningful and consistent across studies. Cut points for low and high dummy groups were created based on either 1 *SD* below and above the norm mean or sample mean (depending on whether measures were standardized), or using the meaningful categories set by the measure creators for the entry skills and behavior measures which contain ratings.

Because different variables were measured within each study, the estimated models also had to differ slightly across studies. Each analysis model predicted the child's school readiness skill using the following generic equation testing the quality main effect:

$$Y_{t+1} = \beta_0 + \beta_1 Y_t + \beta_2 \text{covariates}_t + \beta_3 \text{quality}_{t \to t+1} + e$$

Each potential moderator was examined in separate analyses. The following generic equation tests the potential moderators:

$$\begin{aligned} Y_{t+1} &= \beta_0 + \beta_1 Y_t + \beta_2 \text{covariates}_t + \beta_3 \text{quality}_{t \to t+1} \\ &\quad + \beta_4 \text{moderator}_t + \beta_5 \text{quality}_{t \to t+1} \times \text{moderator}_t + e \end{aligned}$$

In these equations, Y is child school readiness skill at times t (pretest) or $t + 1$ (posttest), covariates_t are covariates from the pretest as described above, $\text{quality}_{t \to t+1}$ is child-care quality averaged over the interval between t and $t + 1$, and moderator_t is the pretest measure of one of the moderators listed above. We estimated several versions of these equations. We then averaged coefficients across the four studies, separately for quality main effects and interactions, in a meta-analysis using the Comprehensive Meta-Analysis software (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Results

Tables 4 through 7 display the standardized regression coefficients and standard errors for the individual studies and the meta-analytic averages for preschool center quality main effects and interactions for the school readiness outcomes in our analyses. A complete listing of the regression results for each individual study can be found in our online appendix.

Results for Language Outcome

Quality Main Effects for Language Outcome

As described above, our analyses began with investigation of each of the four individual studies. These analyses indicated that the ECERS quality score was significantly related to the language outcome in the ECLS–B nationally representative study ($B = .09$, $p < .01$) and the CLASS Instructional Climate quality score was significantly related to the language outcome in the NCEDL 11-state pre-K study ($B = .04$, $p < .05$). The second stage of the analyses combined these quality coefficients from the individual analyses using the meta-analytic approach. As predicted, the meta-analysis revealed a statistically significant, although small, preschool

Table 4
Standardized Coefficients and Meta-Analytic Averages for Preschool Center Quality Main Effects and Interactions for Language Outcome

	NICHD SECCYD	ECLS–B	NCEDL		EHS	**Meta-analytic average**
	ORCE *B* (*SE*)	ECERS *B* (*SE*)	ECERS *B* (*SE*)	CLASS Instructional Climate *B* (*SE*)	ECERS *B* (*SE*)	*B* (*SE*)
M1: **Quality (Q) Main effect**	0.04 (0.03)	0.09** (0.03)	0.03 (0.02)	0.04* (0.02)	0.03 (0.04)	0.05*** (0.01)
Quality × Demographics						
M2: Q × Black	−0.14 (0.09)	−0.03 (0.08)	0.01 (0.03)	0.01 (0.04)	−0.01 (0.07)	−0.01 (0.03)
M2: Q × Hispanic	−0.11 (0.16)	0.05 (0.08)	−0.04 (0.04)	0.03 (0.04)	−0.04 (0.09)	0.00 (0.03)
M2: Q × Other Race or Ethnicity	0.01 (0.12)	–	0.04 (0.04)	0.01 (0.03)	−0.01 (0.16)	0.02 (0.03)
M3: Q × Male	0.01 (0.05)	0.11 (0.06)	0.03 (0.02)	0.02 (0.03)	0.03 (0.06)	0.03 (0.02)
M4: Q × Some College	0.08 (0.06)	−0.04 (0.07)	0.04 (0.03)	0.02 (0.03)	0.03 (0.06)	0.03 (0.02)
M4: Q × BA Plus	0.07 (0.08)	0.14* (0.07)	0.05 (0.03)	0.02 (0.03)	0.29 (0.39)	0.05* (0.03)
Quality × Child's Entry Skills & Behaviors						
M5: Q × Cognitive, Continuous	−0.03 (0.03)	0.02 (0.03)	0.01 (0.01)	0.00 (0.01)	0.02 (0.03)	0.00 (0.01)
M6: Q × Low Cognitive	0.08 (0.08)	−0.05 (0.08)	−0.02 (0.01)	−0.01 (0.01)	−0.11 (0.08)	−0.02 (0.01)
M7: Q × High Cognitive	0.00 (0.06)	0.07 (0.07)	0.00 (0.01)	−0.01 (0.01)	−0.04 (0.07)	0.00 (0.01)
M8: Q × Attention, Continuous	0.01 (0.03)	0.00 (0.03)	0.00 (0.01)	0.01 (0.02)	0.00 (0.03)	0.00 (0.01)
M9: Q × Low Attention	−0.07 (0.07)	−0.02 (0.07)	0.02 (0.01)	0.00 (0.02)	0.05 (0.11)	0.01 (0.01)
M10: Q × High Attention	−0.06 (0.07)	−0.02 (0.09)	−0.01 (0.01)	0.01 (0.01)	0.03 (0.06)	0.01 (0.01)
M11: Q × Problem Beh., Continuous	0.04 (0.02)	–	−0.02 (0.01)	0.00 (0.02)	–	0.01 (0.01)
M12: Q × Most Problem Behaviors	0.11 (0.05)	–	0.02 (0.01)	0.00 (0.01)	–	0.01 (0.01)
M13: Q × Least Problem Behaviors	−0.06 (0.11)	–	−0.01 (0.01)	0.00 (0.01)	–	−0.01 (0.01)

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start; ORCE = Observational Record of the Caregiving Environment; ECERS–R = Early Childhood Environment Rating Scale–Revised; CLASS = Classroom Assessment Scoring System. Reference groups: race or ethnicity = White; gender = female; mother's education = 12 years or less. Models: M1 = quality main effect, pretest, covariates; M2–M13 = adds Quality × Moderator to M1.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

center quality main effects for the language outcome ($B = .05$, $p < .001$). A 1 *SD* increase in quality (roughly 1 point on the 7-point ECERS and CLASS scales and .5 point on the 4-point ORCE scale) was thus associated with a .05 *SD* increase in the language outcome.

Quality Interactions for Language Outcome

Whether observed preschool center quality was moderated was examined by adding the interactions involving the six demographic characteristics in one set of models and the three child entry characteristics in a second set of models (see Table 4 for results). Only one of the six demographic characteristics interacted significantly with quality in the meta-analysis. The Quality × Maternal Education interaction was statistically significant at the meta-analytic level for children's language skills. Preschool center quality was a stronger predictor of language skills for children of highly educated mothers (BA plus) than children whose mothers had 12 years or less of schooling ($B = .05$, $p < .05$), which does not support the compensatory hypothesis but does support the accumulated advantages hypothesis. When the results from each study were examined separately, this interaction attained statistical significance only within the ECLS–B study.

No evidence emerged indicating interactions between preschool center quality and entry skill or behavior levels of the children.

Results for Mathematics Outcome

As shown in Table 5, the meta-analysis also indicated statistically significant, albeit small, preschool center quality main effects for the mathematics outcome ($B = .03$, $p < .05$). A 1 *SD* increase in quality was thus associated with a .03 *SD* increase in the mathematics outcome. Examining the results from each study separately, significant associations involving quality was observed only for the CLASS Instructional Climate main effect in the analysis of the NCEDL data.

The meta-analysis did not yield significant interactions involving observed quality, but a few interactions attained statistical significance within analyses of each study. These are presented, but should not be interpreted in the absence of an overall meta-analytic finding. Regarding Quality × Race or Ethnicity, the CLASS Instructional Climate quality score was a stronger predictor of mathematics achievement for Hispanic and Other (vs. White) children in the NCEDL study. Regarding Quality × Maternal Education, only the ECERS childcare quality measure in the NCEDL was statistically significant—this variable was a stronger predictor for children whose mothers had some college education compared to children whose mothers had 12 or fewer years of education. Finally, an inconsistent scattering of very few statistically significant coefficients with the child entry skills and behaviors yielded nonsignificant findings in the meta-analysis of the mathematics outcome.

Table 5
Standardized Coefficients and Meta-Analytic Averages for Preschool Center Quality Main Effects and Interactions for Mathematics Outcome

	NICHD SECCYD	ECLS–B	NCEDL		EHS	Meta-analytic average
	ORCE *B* (*SE*)	ECERS *B* (*SE*)	ECERS *B* (*SE*)	CLASS Instructional Climate *B* (*SE*)	ECERS *B* (*SE*)	*B* (*SE*)
M1: Quality (Q) Main effect	0.05 (0.03)	0.02 (0.03)	−0.01 (0.02)	0.06*** (0.02)	0.06 (0.04)	0.03* (0.01)
Quality × Demographics						
M2: Q × Black	−0.12 (0.10)	0.01 (0.07)	0.02 (0.03)	0.04 (0.04)	0.01 (0.07)	0.01 (0.03)
M2: Q × Hispanic	−0.15 (0.19)	0.03 (0.09)	0.03 (0.04)	0.14*** (0.04)	−0.01 (0.09)	0.06 (0.03)
M2: Q × Other Race or Ethnicity	−0.17 (0.14)	–	−0.03 (0.04)	0.07* (0.04)	−0.10 (0.15)	0.00 (0.04)
M3: Q × Male	0.02 (0.06)	0.07 (0.06)	0.02 (0.03)	0.01 (0.03)	−0.01 (0.06)	0.02 (0.02)
M4: Q × Some College	0.10 (0.07)	−0.07 (0.07)	0.07* (0.03)	−0.01 (0.03)	0.03 (0.06)	0.03 (0.02)
M4: Q × BA Plus	0.04 (0.09)	0.05 (0.07)	−0.01 (0.04)	−0.07 (0.03)	−0.49 (0.40)	−0.02 (0.03)
Quality × Child's Entry Skills & Behaviors						
M5: Q × Cognitive, Continuous	−0.04 (0.03)	0.00 (0.02)	0.00 (0.02)	−0.01 (0.01)	0.04 (0.03)	0.00 (0.01)
M6: Q × Low Cognitive	0.21* (0.10)	−0.04 (0.07)	−0.01 (0.01)	0.00 (0.02)	−0.21* (0.08)	−0.01 (0.01)
M7: Q × High Cognitive	−0.02 (0.07)	0.00 (0.06)	−0.01 (0.02)	−0.01 (0.01)	−0.06 (0.07)	−0.01 (0.01)
M8: Q × Attention, Continuous	−0.01 (0.03)	0.00 (0.03)	−0.02 (0.02)	0.02 (0.02)	0.01 (0.03)	0.00 (0.01)
M9: Q × Low Attention	0.00 (0.08)	−0.05 (0.07)	0.01 (0.02)	−0.03 (0.02)	0.03 (0.11)	−0.01 (0.02)
M10: Q × High Attention	−0.03 (0.08)	−0.03 (0.09)	0.01 (0.01)	0.02 (0.01)	0.02 (0.06)	0.01 (0.01)
M11: Q × Problem Beh., continuous	0.02 (0.03)	–	0.00 (0.01)	0.04* (0.02)	–	0.02 (0.01)
M12: Q × Most Problem Behaviors	0.09 (0.06)	–	−0.01 (0.02)	−0.04* (0.02)	–	−0.01 (0.02)
M13: Q × Least Problem Behaviors	−0.02 (0.12)	–	−0.01 (0.01)	0.03 (0.02)	–	0.01 (0.02)

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start; ORCE = Observational Record of the Caregiving Environment; ECERS–R = Early Childhood Environment Rating Scale–Revised; CLASS = Classroom Assessment Scoring System. Reference groups: race or ethnicity = White; gender = female; mother's education = 12 years or fewer. Models: M1 = quality main effect, pretest, covariates; M2–M13 = adds Quality × Moderator to M1.
$^{*}p < .05$. $^{***}p < .001$.

Table 6
Standardized Coefficients and Meta-Analytic Averages for Preschool Center Quality Main Effects and Interactions for Social Skills Outcome

	NICHD-SECCYD	ECLS-B	NCEDL		EHS	Meta-analytic average
	ORCE *B* (*SE*)	ECERS *B* (*SE*)	ECERS *B* (*SE*)	CLASS Instructional Climate *B* (*SE*)	ECERS *B* (*SE*)	*B* (*SE*)
M1: **Quality (Q) Main effect**	0.07 (0.04)	0.02 (0.02)	0.00 (0.02)	0.03 (0.02)	n/a (n/a)	0.02 (0.01)
Quality × Demographics						
M2: Q × Black	−0.12 (0.14)	0.04 (0.06)	−0.01 (0.04)	0.00 (0.04)	n/a (n/a)	0.00 (0.03)
M2: Q × Hispanic	−0.09 (0.26)	−0.02 (0.07)	0.02 (0.04)	0.00 (0.04)	n/a (n/a)	0.00 (0.03)
M2: Q × Other Race or Ethnicity	0.09 (0.18)	–	−0.09 (0.05)	−0.02 (0.05)	n/a (n/a)	−0.04 (0.05)
M3: Q × Male	0.03 (0.08)	0.06 (0.04)	0.03 (0.03)	0.03 (0.03)	n/a (n/a)	0.04 (0.02)
M4: Q × Some College	0.05 (0.09)	−0.11* (0.05)	−0.09* (0.04)	−0.04 (0.04)	n/a (n/a)	−0.07* (0.03)
M4: Q × BA Plus	−0.10 (0.12)	0.02 (0.06)	−0.05** (0.04)	−0.07 (0.04)	n/a (n/a)	−0.04 (0.03)
Quality × Child's Entry Skills & Behaviors						
M5: Q × Cognitive, Continuous	−0.04 (0.05)	0.00 (0.02)	−0.05 (0.02)	−0.03 (0.01)	n/a (n/a)	−0.03* (0.01)
M6: Q × Low Cognitive	0.16 (0.12)	−0.02 (0.06)	0.03 (0.02)	0.02 (0.02)	n/a (n/a)	0.02 (0.02)
M7: Q × High Cognitive	0.04 (0.10)	0.02 (0.06)	−0.04* (0.02)	−0.01 (0.02)	n/a (n/a)	−0.02 (0.02)
M8: Q × Attention, Continuous	0.04 (0.04)	0.01 (0.03)	0.00 (0.01)	0.00 (0.02)	n/a (n/a)	0.01 (0.01)
M9: Q × Low Attention	−0.05 (0.10)	0.00 (0.06)	0.02 (0.02)	0.00 (0.02)	n/a (n/a)	0.01 (0.02)
M10: Q × High Attention	0.08 (0.11)	0.00 (0.07)	0.01 (0.01)	0.00 (0.01)	n/a (n/a)	0.01 (0.01)
M11: Q × Problem Beh., Continuous	0.05 (0.03)	–	0.00 (0.02)	−0.02 (0.02)	–	0.01 (0.02)
M12: Q × Most Problem Behaviors	0.08 (0.08)	–	0.00 (0.02)	0.01 (0.02)	–	0.01 (0.02)
M13: Q × Least Problem Behaviors	−0.20 (0.15)	–	0.00 (0.02)	−0.02 (0.02)	–	−0.01 (0.02)

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start; ORCE = Observational Record of the Caregiving Environment; ECERS–R = Early Childhood Environment Rating Scale–Revised; CLASS = Classroom Assessment Scoring System. Reference groups: race or ethnicity = White; gender = female; mother's education = 12 years or fewer. Models: M1 = quality main effect, pretest, covariates; M2–M13 = adds Quality × Moderator to M1.
**p* < .05. ***p* < .01.

Results for Social Skills Outcome

As shown in Table 6, the meta-analyses of the social skills outcome did not indicate that preschool center quality was related to gains in social skills in preschool-age center care, but did reveal significant interactions for one of six interactions between quality and demographic and one of three interactions between quality and child entry skills. The two statistically significant findings with social skills outcomes were with maternal education and the entry cognitive skill (as a continuous variable) moderators. First, our results suggest that preschool center quality was a stronger predictor of gains in social skills for children whose mothers had a high school degree or less than for children whose mothers had some college ($B = -0.07$, $p < .05$), which provides limited support for the compensatory hypothesis. Second, with regard to the Quality × Cognitive interaction, analyses that included entry-level cognitive skills as a continuous variable suggested that that child-care quality was a stronger predictor of gains in social skills among children who entered preschool-age center care with lower cognitive skills ($B = -.03$, $p < .05$). Failure to attain this interaction when it was tested using categorical measures of entry cognitive skills is not surprising given the very modest interaction term that was attained in the analyses involving the continuous measure of entry cognitive skills and the con-

Table 7

Standardized Coefficients and Meta-Analytic Averages for Preschool Center Quality Main Effects and Interactions for Problem Behaviors Outcome

	NICHD-SECCYD	ECLS–B	NCEDL		EHS	Meta-analytic average
	ORCE *B (SE)*	ECERS *B (SE)*	ECERS *B (SE)*	CLASS Instructional Climate *B (SE)*	ECERS *B (SE)*	*B (SE)*
M1: **Quality (Q) Main effect**	0.01 (0.04)	−0.03 (0.03)	0.01 (0.02)	0.05** (0.02)	0.00 (0.05)	0.01 (0.01)
Quality × Demographics						
M2: Q × Black	0.13 (0.14)	−0.01 (0.07)	0.03 (0.04)	0.02 (0.05)	0.03 (0.08)	0.02 (0.03)
M2: Q × Hispanic	0.28 (0.24)	0.02 (0.09)	0.00 (0.04)	0.01 (0.04)	0.11 (0.09)	0.03 (0.03)
M2: Q × Other Race or Ethnicity	−0.15 (0.17)	–	0.01 (0.05)	0.04 (0.06)	0.09 (0.19)	0.01 (0.05)
M3: Q × Male	0.01 (0.08)	−0.08 (0.05)	0.01 (0.03)	0.02 (0.03)	0.01 (0.06)	−0.01 (0.02)
M4: Q × Some College	0.01 (0.10)	0.12* (0.06)	0.01 (0.04)	0.06 (0.04)	−0.02 (0.07)	0.04 (0.03)
M4: Q × BA Plus	0.02 (0.12)	−0.02 (0.06)	−0.04 (0.04)	−0.10** (0.04)	0.40 (0.47)	−0.05 (0.03)
Quality × Child's Entry Skills & Behaviors						
M5: Q × Cognitive, Continuous	0.01 (0.04)	0.00 (0.02)	−0.02 (0.02)	−0.02 (0.02)	−0.05 (0.03)	−0.01 (0.01)
M6: Q × Low Cognitive	0.09 (0.12)	0.01 (0.07)	0.03 (0.02)	0.03 (0.02)	0.07 (0.10)	0.03 (0.02)
M7: Q × High Cognitive	0.04 (0.10)	−0.03 (0.07)	−0.02 (0.02)	0.00 (0.02)	−0.10 (0.07)	−0.02 (0.02)
M8: Q × Attention, Continuous	0.02 (0.04)	0.00 (0.03)	0.01 (0.02)	0.00 (0.02)	−0.05 (0.03)	−0.01 (0.01)
M9: Q × Low Attention	−0.09 (0.10)	−0.01 (0.06)	−0.01 (0.02)	−0.02 (0.02)	0.18 (0.13)	−0.01 (0.02)
M10: Q × High Attention	−0.01 (0.10)	0.02 (0.07)	0.01 (0.01)	−0.01 (0.01)	−0.04 (0.07)	0.00 (0.01)
M11: Q × Problem Beh., Continuous	−0.06 (0.03)	–	0.00 (0.02)	−0.03 (0.02)	–	−0.03 (0.02)
M12: Q × Most Problem Behaviors	−0.11 (0.08)	–	0.01 (0.02)	0.03 (0.03)	–	0.01 (0.02)
M13: Q × Least Problem Behaviors	0.29 (0.15)	–	0.01 (0.02)	−0.02 (0.02)	–	0.00 (0.02)

Note. NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; NCEDL = National Center for Early Development and Learning; EHS = Early Head Start; ORCE = Observational Record of the Caregiving Environment; ECERS–R = Early Childhood Environment Rating Scale–Revised; CLASS = Classroom Assessment Scoring System. Reference groups: race or ethnicity = White; gender = female; mother's education = 12 years or less. Models; M1 = quality main effect, pretest, covariates; M2–M13 = adds Quality × Moderator to M1.
*$p < .05$. **$p < .01$.

sistency in the direction of effects for the analyses involving the categorical variables.

Results for Problem Behaviors Outcome

At the meta-analytic level, no evidence emerged indicating that preschool center quality was related to externalizing problem behaviors as a main or interaction effect involving either demographic characteristics or entry child skills or behaviors (see Table 7). Again, we examined the results from separate analyses of each study to see if there was any suggestion of associations. In the NCEDL data, higher quality was related to more problem behaviors ($B = .05$, $p < .01$), but the Quality × Maternal Education interaction suggested that this positive association was stronger for children whose mothers had a high school degree or less than among children whose mothers had a BA degree or more ($B = -.10$, $p < .01$). In contrast, in the ECLS–B data, higher quality was related to fewer problem behavior among children whose mothers had a high school or less than among children whose mothers had some college ($B = .12$, $p < .05$). Failure to attain significant quality effects in the meta-analysis and the small, somewhat contradictory findings in the individual studies provide little evidence that preschool center quality was related to change in problem behaviors in these studies.

Discussion

The purpose of this study was to examine associations between child-care quality in preschool center care and children's academic and socioemotional skills at kindergarten entry. Our study involved a comprehensive approach to ensure that the same analytic strategy was used across selected large child-care studies and then meta-analytically combining the findings across studies. Our focus is on child-care quality as a predictor of gains in language, mathematics, social skills, and externalizing problem behaviors during preschool; therefore, we are looking at the impact of the quality of preschool center care for 1–2 years—not the quality of care for the entire early childhood period. Our effect sizes are like value-added estimates and do not address questions about the impact of prior quality of care or of quality of care that is not measured by these widely used quality measures.

Results from these analyses suggest that observed quality of the preschool center classroom is very modestly related to: acquisition of language and mathematics skills overall, language skills for children of highly educated mothers (BA plus), and social skills for children who entered preschool-age care with lower cognitive skills or had mothers with some college. The quality main effects in this study are similar to those reported in several recent secondary analyses of the child-care literature and the meta-analysis of published articles examining child-care quality in preschool center classrooms and child outcomes conducted by Burchinal et al. (2011).

The magnitudes of these associations are considerably smaller than those reported by advocates. Several factors contribute to these differences. This study examined effects of quality of preschool care over a 1-year period, controlling for prior functioning. In some cases, the amount of exposure is 4–5 months, a time period that is much smaller than those reported when child-care quality is measured across a longer span of the early childhood period and also much smaller than those attained from carefully controlled small experimental studies.

Also, it is possible that child-care quality may not be adequately measured in currently available studies (Burchinal et al., 2011; Farran & Hofer, 2011; Gormley, 2007; Zaslow et al., 2006). All of the present quality measures were developed on a conceptual basis by child development experts without detailed psychometric analysis. For example, the ORCE quality measure developed in the 1990s specifically for the NICHD SECCYD was originally designed to focus on quality of interactions and relationships with teachers rather than cognitive and achievement skills. The end of the 1990s saw a paradigm shift in this research area as child care began being viewed as a compensatory intervention for at-risk children rather than as a service primarily for middle-class children and families. As the focus shifted to child-care's ability to improve school entry academic skills for at-risk students, the focus on instructional practices increased, yet none of the quality measures included intentional instructional practices (such as verbal interactions, which include elaborated vocabulary and sentence structures) that are important for fostering children's cognitive, academic, and social development (Burchinal et al., 2011). Furthermore, none measured the quality of instruction in specific content areas (e.g., language). The experimental evaluations of successful curricula suggest that a combination of the behaviors measured by these quality scales (e.g., warmth and responsiveness) as well as those not measured by these scales (e.g., focus and depth of instruction within the content area) is needed to positively affect children's school readiness skills.

Indeed, in a recent psychometric analysis of the ECERS–R, researchers advise the child-care quality field to proceed cautiously in using the ECERS–R to assess child-care quality as they find little evidence of response process validity, structural validity, and criterion validity (Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2012). Also, given that the ECERS emphasizes aspects of the environment (e.g., materials, health practices, safety, and discipline) as opposed to educational goals (e.g., language, literacy, and numeracy), the ECERS may not be well suited for school readiness studies (Zaslow et al., 2006). The CLASS Instructional Climate measure—the most recent and most instruction-focused of the three quality measures in our study—yielded the most quality main effects across the four outcomes we examined, which contributed to the statistically significant meta-analytic average for the language and mathematics outcomes. On the whole, however, our findings are consistent with other recent studies suggesting that currently available quality measures may not be adequate to the research tasks being undertaken (Burchinal et al., 2011; Gordon et al., 2012; Zaslow et al., 2006).

We are cautious about overinterpreting the evidence of compensatory or accumulated advantages and skill begets skill as there was very little evidence supporting either hypothesis. In our examination of six demographic characteristics and three child entry skills and behaviors, the language outcome indicated only one statistically significant interaction while the social skill outcome indicated only two statistically significant interactions, with no evidence of moderating effects for the mathematics or problem behavior outcomes at the meta-analytic level. Failure to observe convincing evidence for either hypothesis could reflect a lack of either a compensatory or Matthew effect in preschool-age child-care impacts on school readiness outcomes, or it could reflect lack of power to detect very modest differences. If the quality measures provide relatively weak measurement of classroom quality, then it will be even more difficult to detect interactions using those measures than it would be to detect main effects due to the decreased power inherent in testing interactions (Cohen, 1988). Replication of the attained interactions in other work will increase confidence that the findings are robust.

These findings may have implications for recent policies attempting to improve quality of child care through greater focus on observed quality ratings. Head Start has used the ECERS–R as part of its evaluation study, FACES, and is switching to the CLASS in performance monitoring. Soon, programs with low scores on the CLASS will be given greater scrutiny in the reauthorization process. Over half the states are using quality rating scales, typically the ECERS, in the quality rating and improvement systems based on evidence linking observed quality to child outcomes (Tout et al., 2010).

Study Limitations

This study has several limitations. First, as discussed earlier, although we employed the most widely used child-care quality measures, they may not provide adequate assessment of all aspects of the preschool center classroom that impact children's school readiness skills.

Second, because covariates varied somewhat across the four studies, our regression models, while similar to one another, were not identical across studies. For example, an important difference is the variation in child's age at assessment and the length interval between pretest (i.e., lagged dependent variable) and outcome for the four studies. We attempted to account for this variation by including controls for *age at pretest assessment* and the *age at outcome–age at assessment* time interval, but differences across studies remain.

Third, we only examined preschool-aged children (3–5 years old) in center-based child-care settings so our results should not be generalized to children who did not select into this context. We would, however, point out that over 75% of children nationally and within almost all early child-care studies are in center-based care at 4–5 years of age (Barnett, Epstein, Friedman, Sansanelli, & Hustedt, 2009).

Fourth, only the nationally representative ECLS–B offers extensive diversity in terms of the selected demographic characteristics and child entry skills and behaviors used to test interactions between observed quality and child outcomes. The other three studies tended to be more homogeneous, with two studies that include primarily low-income families (NCEDL and EHS) and one study that includes primarily middle-class families (NICHD SECCYD). We acknowledge that there are confounds in the NICHD SECCYD study, particularly between race or ethnicity and poverty status.

Finally, the three process quality measures in this study are different from one another. The ECERS and CLASS scores tend to be moderately to strongly correlated (Pianta et al., 2008), but it is likely that differences in these measures could reduce the likelihood of consistent findings across studies. We nevertheless believe that there is value in including them in our analysis as they are well established and frequently used measures in the early child-care quality literature. The purpose of employing meta-analysis techniques is to combine estimates across studies that have examined the same issue and to ask about the strength of the evidence across a diverse set of studies that include some of the major child-care studies of the last 15 years. We felt it was better to address our important questions in a systematic way even though we acknowledge that the study has limitations.

Conclusion

Our article applies meta-analytic techniques to summarize results from parallel analyses of four recent child-care studies to estimate the following: (a) the magnitude of the association between preschool center quality and child outcomes, and (b) variation in quality impacts by demographic characteristics and child entry skills and behaviors. The consistency of our results and the precision with which they are estimated across the four studies, multiple outcomes, and multiple child-care quality measures suggest the following: (a) there are, at best, very small preschool center quality main effects on child language and mathematics outcomes, and (b) there is generally an absence of *differential* preschool center quality effects on school readiness for subgroups of children defined by demographic characteristics or child entry skills and behaviors.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont Department of Psychiatry.

Allison, P. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Anderson, M. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. *Journal of the American Statistical Association, 103,* 1481–1495. doi:10.1198/016214508000000841

Barnett, W. S., Epstein, D. J., Friedman, A. H., Sansanelli, R. A., & Hustedt, J. T. (2009). *The state of preschool 2009: State preschool yearbook.* New Brunswick, NJ: National Institute for Early Education Research, Rutgers Graduate School of Education.

Bassok, D. (2010). Do Black and Hispanic children benefit more from preschool? Understanding differences in preschool effects across racial groups. *Child Development, 81,* 1828–1845. doi:10.1111/j.1467-8624.2010.01513.x

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: Psychological Corporation.

Belfield, C., Nores, M., Barnett, W. S., & Schweinhart, L. (2005). Updating the benefit-cost analysis of the High/Scope Perry Preschool Program through age 40. *Educational Evaluation and Policy Analysis, 27,* 245–262.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

Bronfenbrenner, U., & Morris, P. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 793–828). New York: Wiley.

Brooks-Gunn, J., Han, W., & Waldfogel, J. (2002). Maternal employment and child cognitive outcomes in the first three years of life: The NICHD Study of Early Child Care. *Child Development, 73,* 1052–1072. doi:10.1111/1467-8624.00457

Burchinal, M. R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly, 18,* 401–426. doi:10.1016/j.ecresq.2003.09.003

Burchinal, M. R., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Development Science, 12,* 140–153. doi:10.1080/10888690802199418

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow (Ed.), *Quality measurement in early childhood settings.* Baltimore, MD: Brooks.

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science, 4,* 149–165.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2009). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25,* 166–176. doi:10.1016/j.ecresq.2009.10.004

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112,* 579–620.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37,* 231–242. doi:10.1037//0012-1649.37.2.231

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Erlbaum.

Cunha, F., Heckman, J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence of life cycle skill formation. In F. Hanushek & F. Welch (Eds.), *Handbook of the economics of education.* North Holland, Amsterdam: Elsevier.

Dearing, E., McCartney, K., & Taylor, B. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development, 80,* 1329–1349. doi:10.1111/j.1467-8624.2009.01336.x

DeCoster, J., Iselin, A.-M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods, 14,* 349–366. doi:10.1037/a0016956

Duncan, G., Dowsett, C., Classens, A., Magnuson, K., Huston, A., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446. doi:10.1037/0012-1649.43.6.1428

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.) Circle Pines, MN: American Guidance Service.

Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., . . . Barnett, W. S. (2005). *Pre-kindergarten in eleven states: NCEDL's Multi-State Study of Pre-Kindergarten and study of State-Wide Early Education Programs (SWEEP): Preliminary descriptive report.* Chapel Hill, NC: National Center for Early Development & Learning. Retrieved June 1, 2010, from http://www.fpg.unc.edu/%7encedl/pdfs/SWEEP_MS_summary_final.pdf

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., . . . Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78,* 558–580. doi:10.1111/j.1467-8624.2007.01014.x

Elder, G. H., Jr (1996). Human lives in changing societies: Life course and developmental insights. In R. B. Cairns, G. H. Elder Jr., & E. J. Costello (Eds.), *Developmental science* (pp. 31–62). New York: Cambridge University Press.

Farran, D. C., & Hofer, K. (2011). Evaluating the quality of early childhood education programs. In O. Saracho & B. Spodek (Eds.), *Handbook of research on the education of young children*. New York: Routledge/Taylor & Francis.

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2012). An assessment of the validity of the ECERS–R with implications for measures of child care quality and relations to child development. *Developmental Psychology*. Advance online publication. doi:10.1037/a0027899

Gormley, W. T. (2007). Early childhood care and education: Lessons and puzzles. *Journal of Policy Analysis and Management*, *26*, 633–671. doi:10.1002/pam.20269

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, *41*, 872–884. doi:10.1037/0012-1649.41.6.872

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale* (Rev ed.) New York: Teachers College Press.

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., et al. (1986). The Teacher-Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review*, *15*, 393-409.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D. M., & Clifford, R. (2008). Ready to learn? children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*, 27–50. doi:10.1016/j.ecresq.2007.05.002

Kagan, S.L., Moore, E., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Washington, DC: National Education Goals Panel, Goal 1 Technical Planning Group.

Knudsen, E., Heckman, J., Cameron, J., & Shonkoff, J. (2006). Economic, neurobiological and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 10155–10162.

Lamb, M. E., & Ahnert, L. (2006). Nonparental child care: Context, concepts, correlates, and consequences. In W. Damon & R. M. Lerner (Series Eds.) & K. A. Renninger & I. E. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 950–1016). New York: Wiley.

Levine, S., Elzey, F. F., & Lewis, M. (1969). *California Preschool Social Competency Scale*. Palo Alto, CA: Consulting Psychologists Press.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66. doi:10.1016/j.econedurev.2005.11.005

Love, J. M., Harrison, L., Sagi-Schwartz, A., van Ilzendoorn, M. H., Ross, C., Ungerer, J. A., et al. (2003). Child care quality matters: How conclusions may vary with context. *Child Development*, *74*, 1021–1033. doi:10.1111/1467-8624.00584

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., . . . Vogel, C. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, *41*, 885–901. doi:10.1037/0012-1649.41.6.885

Magnuson, K., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*, 33–51. doi:10.1016/j.econedurev.2005.09.008

Mashburn, A. J., Pianta, R. C., Hamre, B., Downer, J., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*, 732–749. doi:10.1111/j.1467-8624.2008.01154.x

Matthews, J., Ponitz, C., & Morrison, F. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, *101*, 689–704. doi:10.1037/a0014240

NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, *71*, 960–980. doi:10.1111/1467-8624.00202

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*, 133–164.

NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology*, *39*, 451–469. doi:10.1037/0012-1649.39.3.451

NICHD Early Child Care Research Network. (2006). Child care effect sizes for the NICHD Study of Early Child Care and youth development. *American Psychologist*, *61*, 99–116. doi:10.1037/0003-066X.61.2.99

NICHD Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*, 1454–1475. doi:10.1111/1467-8624.00617

Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly*, *43*, 451–477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*, 1534–1553. doi:10.1111/1467-8624.00364

Pianta, R. C., Burchinal, M., Barnett, E. S., & Thornburg, K. (2009). Preschool in the United States: What we know, what we need to know, and implications for policy and research. *Psychological Science in the Public Interest*, *10*, 49–88. doi:10.1177/1529100610381908

Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System (CLASS) manual, Prek*. Baltimore, MD: Brookes.

Reynolds, A., Temple, J., Robertson, D., & Mann, E. (2002). Age 21 cost-benefit analysis of the Title I Chicago child-parent centers. *Educational Evaluation and Policy Analysis*, *24*, 267–303.

Sameroff, A. J., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaker casualty. In F. D. Horowitz (Ed.), *Review of child development research* (Vol. 4, pp. 187–244). Chicago: University of Chicago Press.

Sapolsky, R. (2004). Mothering style and methylation. *Nature Neuroscience*, *7*, 791-792.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.

Snow, K., Derecho, A., Wheeless, S., Lennon, J., Rosen, J., Rogers, J., et al. (2009). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), kindergarten 2006 and 2007 data file user's manual (2010–010)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *The Child Care Quality Rating System (QRS) assessment: Compendium of quality rating systems and evaluations*. Washington, DC: Child Trends.

Vandell, D. L. (2007). Early child care: The known and the unknown. In G. Ladd (Ed.), *Appraising the human development sciences essays in honor of Merrill-Palmer Quarterly* (pp. 300–328). Detroit, MI: Wayne State University Press. (Reprinted from Merrill-Palmer Quarterly, 50, 387-414.)

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N., & NICHD Early Child Care Research Network. (2010). Do effects of early child care extend to age 15 years? *Child Development*, *81*, 737–756. doi:10.1111/j.1467-8624.2010.01431.x

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, *75*, 296–312. doi:10.1111/j.1467-8624.2004.00670.x

Watamura, S. E., Phillips, D. A., Morrissey, T. W., McCartney, K., & Bub, K. (2011). Double jeopardy: Poorer social-emotional outcomes for children in the NICHD SECCYD experiencing home and child care environments that confer risk. *Child Development*, *82*, 48–65. doi:10.1111/j.1467-8624.2010.01540.x

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Itasca, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III: Tests of achievement*. Itasca, IL: Riverside..

Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer, L., et al. (2006). Child outcome measures in the study of child care quality. *Evaluation Review*, *30*, 577–610. doi:10.1177/0193841X06291529

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool Language Scale-3*. San Antonio, TX: Psychological Corporation.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Dataset-Specific CONTROLS.

Appendix S2. Regression Tables for the Individual Datasets.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1191–1208

Child Care in Infancy and Cognitive Performance Until Middle Childhood in the Millennium Cohort Study

Sylvana M. Côté
University of Montreal

Orla Doyle
University College Dublin

Amélie Petitclerc
Columbia University

Lori Timmins
University of British Columbia

This study used a British cohort ($n = \sim 13{,}000$) to investigate the association between child care during infancy and later cognition while controlling for social selection and missing data. It was found that attending child care (informal or center based) at 9 months was positively associated with cognitive outcomes at age 3 years, but only for children of mothers with low education. These effects did not persist to ages 5 or 7 years. Early center-based care was associated with better cognitive outcomes than informal care at ages 3 and 5 years, but not at 7 years. Effect sizes were larger among children whose mother had low education. Propensity score matching and multiple imputation revealed significant findings undetected using regression and complete-case approaches.

The majority of children in developed countries receive nonmaternal child-care services during early childhood despite widespread concern about their potentially detrimental impact on children's social and cognitive development (UNICEF, 2008). Furthermore, in recent years, many governments have adopted the explicit goal of expanding child-care coverage, often through publicly funded universal child-care subsidies. Research examining the impact of child care on child developmental outcomes has led to substantially heterogeneous results, and the direction and magnitude of the effects vary according to several important parameters (NICHD ECCRN, 2006). These parameters include the age at which child care is initiated, the use of child care or maternal employment as a predictor of child development, the type of child care under consideration (e.g., formal vs. informal), and whether family characteristics moderate the associations between child care and cognitive outcomes.

In this study, we examine the impact of child care on children's cognitive outcomes at 3, 5, and 7 years with one of the largest data sets available, the British Millennium Cohort Study ($n = \sim 13{,}000$). Note that the term *impact* is used throughout to simplify the text and does not imply we are drawing causal interpretations from the data. Given the mixed findings in the literature and the controversies surrounding the impact of the type of child-care services received in infancy, our aim is to use this unique study to examine the role of formal and informal child-care services received in the first 9 months of life and to account for two methodological challenges inherent in longitudinal child-care research: social selection (using propensity score matching [PSM]) and missing data (using multiple imputation [MI]).

Child Care and Children's Cognitive Development: Sources of Variations

Concerns about the impact of nonmaternal care early in life are rooted in the human and nonhuman attachment literatures, which suggest that maternal proximity and availability during infancy are essential to the offspring's healthy development (Belsky, 2001; Bowlby, 1951). Two bodies of literature have quantified the impact of maternal unavailability on child development. In the psychological literature, maternal unavailability is typically associated with the use of child-care services, whereas in the

---

Permission to use the U.K. Millennium Cohort Study given by the ESRC Data Archive at Essex is gratefully acknowledged.

Correspondence concerning this article should be addressed to Sylvana Côté, GRIP, University of Montreal, 3050 Edouard-Montpetit, Montréal, QC H3T 1J7. Electronic mail may be sent to sylvana.cote@gmail.com.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0007
DOI: 10.1111/cdev.12049

economic literature it is associated with maternal employment outside the home (cf. Burchinal & Clarke-Stewart, 2007). Child-care use and maternal employment are only partially overlapping constructs. That is, while the majority of mothers who use child care are in the labor market, a minority are not. In addition, some working mothers may use nonparental child-care services by working night shifts or working from home for example.

Studies examining the role of child-care services on cognitive development generally find positive effects (Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; NICHD ECCRN, 2000a, 2000b, 2002), although these are often limited to certain types of child-care services or to certain age groups. Specifically, positive impacts were reported for center-based care or formal child-care services and for services initiated in toddlerhood (Hansen & Hawkes, 2009; Loeb et al., 2007; NICHD ECCRN, 2004). Studies that have focused on the type of child care in infancy and later cognitive development find mixed results, with some reporting positive effects (Andersson, 1989; Hansen & Hawkes, 2009; Sylva, Stein, Leach, Barnes, & Malmberg, 2011) and others reporting no effect (e.g., Jaffee, Van Hulle, & Rodgers, 2011; Park & Honig, 1991). Studies on the association between maternal employment and cognitive development often report negative effects of early maternal work (e.g., Baydar & Brooks-Gunn, 1991; Berger, Brooks-Gunn, Paxson, & Waldfogel, 2008; Brooks-Gunn, Han & Waldfogel, 2002; Han, Waldfogel, & Brooks-Gunn, 2001; Ruhm, 2004; Waldfogel, Han, & Brooks-Gunn, 2002). These results are particularly interesting given that maternal employment is generally associated with higher family income, which facilitates greater investment in child inputs. The negative effects are usually detected for maternal employment in the 1st year postbirth and are more pronounced for full-time employment than part-time employment (Baydar & Brooks-Gunn, 1991; Brooks-Gunn et al., 2002; Hill, Waldfogel, Brooks-Gunn, & Han, 2005; Waldfogel et al., 2002).

One potential source of divergence between the child-care and maternal employment literature is the lack of consideration, particularly in many maternal employment studies, of the type of care the child received. One study that did examine the role of child-care services among children of working mothers found a positive effect for center-based services and negative effects for care by a friend or relative (Gregg, Washbrook, Propper, & Burgess, 2005). Formal care is recognized as being, on average, more structured and of better quality than informal care and previous studies indicate the importance of considering the type of care (e.g., Gregg et al., 2005; Sylva et al., 2011).

Variations by Family Risk

There is evidence to suggest that the impact of child care or maternal employment may vary as a function of family socioeconomic status (SES; Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006; Côté, Borge, Geoffroy, Rutter, & Tremblay, 2008; Côté et al., 2007; Geoffroy et al., 2007; Votruba-Drzal, Coley, & Chase-Lansdale, 2004). Two main variables have been used to reflect socioeconomic risk—low maternal education and low family income—and three moderating hypotheses have been proposed. Each is reviewed next.

First, there is support for the "compensatory hypothesis," which suggests that child care may be particularly beneficial for low-SES children as it provides cognitive stimulation and opportunities for learning that may not be available in the home. Consistent with this hypothesis, Geoffroy et al. (2010) showed that children of mothers with low levels of education had scores on school readiness and achievement tests at 6 and 7 years that did not differ from the children of high-educated mothers if they received formal child care. The same compensatory effect was found using either low income (Caughy, DiPietro, & Strobino, 1994) or a composite index including low income, low education, and low occupational prestige as moderators (Geoffroy et al., 2007). Second, there is support for the "loss of resource hypothesis," which suggests that children from middle- to high-SES families may be deprived of the stimulation and learning opportunities provided by their families in situations of nonmaternal care. The loss of resource hypothesis was supported in a child-care study showing that among families providing stimulating home environments, children receiving center-based child care in infancy had lower mathematics scores (Caughy et al., 1994). In addition, maternal employment studies reported negative effects of maternal employment among families of mothers with high levels of education (Brooks-Gunn et al., 2002; Gregg et al., 2005; Han et al., 2001; Harvey, 1999; Hill et al., 2005; Ruhm, 2004). Third, there is some support for the "dual risk hypothesis" suggesting that among children from already impoverished backgrounds, maternal employment may represent an additional risk. For example, Waldfogel et al. (2002) reported negative effects of full-time maternal employment during the 1st year of life on low-income children's Peabody Picture Vocabulary Test–Revised and Peabody Individual Achievement Test scores at 4 years. Finally, some studies found no or little evidence of an interaction between sociodemographic characteristics of the family and child care, but rather a main effect of high-quality care on child outcomes (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; NICHD & Duncan, 2003; NICHD ECCRN, 2002).

Given these mixed findings, it is important to systematically test whether children from disadvantaged families benefit more or less from child care than those from nondisadvantaged families. Low income and low maternal education are the most frequently used indicators of disadvantage, yet surprisingly few studies have compared the moderating role of one versus the other. Maternal education and family income are highly correlated but not completely overlapping constructs, and there are good reasons for testing both. For example, from a policy perspective maternal education is a useful indicator that can be used to target populations as it is one of the strongest predictors of children's cognitive development (Peisner-Feinberg et al., 2001; Sameroff, Seifer, Baldwin, & Baldwin, 1993) and can be easily measured. Low family income, although harder to accurately assess than education, is also correlated with family adversity (Caughy et al., 1994; Dearing, McCartney, & Taylor, 2009) and has often been used to target populations for service delivery purposes. One objective of this study is to test the interactions between child care and both maternal education and low income to compare their potentially moderating roles.

Accounting for Social Selection

A major conceptual and methodological issue in child-care research is the possibility that the selection of families into child-care services explains part of the association with children's outcomes. Child-care studies typically control for selection characteristics in ordinary least squares (OLS) regression models. These characteristics generally include parenting and parental cognitive abilities, levels of stimulation in the home, and income. In addition, family characteristics such as ethnicity (Early & Burchinal, 2001; NICHD ECCRN, 1997), whether the child has other siblings (Singer, Fuller, Keiley, & Wolf, 1998), and maternal beliefs about the effects of maternal employment (NICHD ECCRN, 1997) have been suggested as potential selection factors. The decision to use child care may also depend on children's characteristics such as temperament (Lerner & Galambos, 1986). However, it is rarely possible to account for all potential mediating factors.

Two maternal employment studies have used PSM (Berger et al., 2008; Hill et al., 2005), which is a method designed to reduce selection bias by matching individuals who received a treatment (e.g., child care) with individuals who did not (e.g., maternal care), but have similar probabilities of being treated based on measured characteristics (Dehejia & Wahba, 2002; Rosenbaum & Rubin, 1983; Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007). They reported minimal differences between the regression and PSM approaches (Berger et al., 2008; Hill et al., 2005). Apart from one study examining the impact of participation in Head Start (Zhai, Brooks-Gunn, & Waldfogel, 2011), we did not find any published studies using PSM to examine the association between child-care services during infancy and cognitive development. Given that the selection factors for child-care use and maternal employment may differ, child-care studies accounting for social selection with PSM may enrich the debate. One aim of this study is to compare the results obtained with OLS regression and PSM approaches to account for selection biases.

Accounting for Attrition

Attrition is inherent in longitudinal studies and may create a missing value problem as participants lost to follow-up usually differ from those who remain in the study. One strategy to reduce attrition bias is MI, that is, to fill in missing values with multiple predictions based on observed data. There is evidence that MI, as compared to a complete-case approach, can yield different results. For instance, Hill et al. (2005) found that MI reduced the negative impact of maternal employment by supplementing the subgroup in which there was no negative effect—the group of disadvantaged families (Hill et al., 2005). This underlies the importance of carefully accounting for attrition in studies on the long-term impact of child care.

The Setting

The majority of studies on child care and cognitive development have been conducted in the United States. As results may differ in settings with different maternal leave and child-care policies, and where the provision, users, and quality of child-care services differ, it is important to examine whether the pattern of findings differ in other countries. The United States is unique among Western countries in relation to national parental leave policies. Although a federal policy now mandates a 12-week, job-protected maternity leave, there is no mandatory payment at the national level. In contrast, paid maternal leave is provided in the United Kingdom and is of longer duration. When children in this study were born, mothers in the United Kingdom were allowed 18 weeks of paid leave, and those with 1 year of continuous service could take an additional 13 weeks of unpaid parental leave during the child's first 5 years (Waldfogel, 2001). Not surprisingly, return to work patterns differ markedly across the two countries (Crosby & Hawkes, 2007). In the United Kingdom, approximately 8% of new mothers are back to work within 3 months, whereas the equivalent is 40% in the United States (Berger & Waldfogel, 2004). Furthermore, 70% of U.K. mothers who return by 6 months are employed part-time compared to only 40% in the United States.

Both the U.K. and U.S. governments provide assistance to families using child care in the form of income tax credits and targeted assistance for lower income families. Public expenditures on child care are about 4 times higher in the United Kingdom than the United States as a function of gross domestic product or per child aged younger than 3 years (Organisation for Economic Co-operation and Development [OECD], 2011). Nevertheless, as the cost of formal child care is much higher in the United Kingdom, the net cost of formal child care, as a function of families' income, remains higher in the United Kingdom than in the United States (OECD, 2011). These different policies and the differential patterns of working practices and child-care use point to the need for studies from multiple countries.

Study Objectives

Using data from the British Millennium Cohort Study (MCS), a prospective longitudinal survey of children born between 2000 and 2002, we address three substantive questions and two methodological questions: (a) As compared to parental care, is any form of child care (informal or center based) during infancy associated with better or poorer short-term (age 3 years) and medium-term (5 and 7 years) cognitive outcomes? (b) As compared to informal child care (e.g., care by a family member, neighbor), is center-based child care in infancy associated with better or poorer cognitive outcomes? (c) Is there a differential association between child care and cognitive development as a function of the type of socioeconomic risk (i.e., low maternal education or low family income)? (d) Does accounting for selection biases using PSM yield results similar to when using OLS regression? (e) Does accounting for attrition with MI affect the results?

Method

Sample and Procedures

The MCS is a nationally representative sample of 18,819 children born in the United Kingdom between September 2000 and January 2002. The sample was clustered geographically by electoral wards and constructed to overrepresent areas of disadvantage, communities with high concentrations of ethnic minorities, and the three smaller countries of the United Kingdom (i.e., Northern Ireland, Wales, and Scotland). The sample was identified through Child Benefit records provided by the Department of Social Security (for more information, see Dex, 2005). The overall response rate was 72% (Plewis, 2004). This study utilizes the first four waves of the survey conducted when the children were 9 months, and 3, 5, and 7 years old.

Dependent Variables: Children's Cognitive Skills at Ages 3, 5, and 7 Years

This study employed two measures of cognitive development at age 3: the Bracken Basic Concept Scale–Revised (BBCS–R) and the British Ability Scale (BAS); two measures at 5 years: the BAS and the Foundation Stage Profiles (FSP); and one measure at 7 years: the BAS. All assessments, apart from the FSP, were conducted in the home by a trained interviewer. The FSP was completed by teachers. For comparison purposes all cognitive outcomes were standardized to have a mean of 100 and a standard deviation of 15.

The BBCS–R (age 3 years). The BBCS–R is a test of academic skills suitable for children aged 2½ and 8 years (Bracken, 1984). It has been used to predict future academic achievement (Panter, 2000). The results of the pilot study and a consultation process led MCS to administer six subtests of the BBCS–R. These tests incorporated 88 concepts of children's knowledge of colors, letter identification, numbers or counting, sizes, comparisons, and shape recognition. The BBCS–R score was available for 13,651 children at age 3.

The BAS II (at 3, 5, 7 years). The BAS measures cognitive abilities and educational achievements among children between the ages 2½ and 8 years (Elliott, Smith, & McCulloch, 1996, 1997). Age-specific combinations of scales were used at different ages. At 3 years, the BAS Naming Vocabulary scale was used. At 5 years, the BAS Naming Vocabulary, Picture Similarity, and Pattern Construction scales were used. There was a moderate intercorrelation between the three domains ranging from $r = .36$ ($p < .001$) to $r = .37$ ($p < .001$). At 7 years, the BAS Reading, Pattern Construction, and Number Skills scales were used. Again, there was moderate to high intercorrelation between the three domains ranging from $r = .33$ ($p < .001$) to $r = .52$ ($p < 0.001$). For the ages 5 and 7 BAS measures, two overall BAS scores were generated using the mean scores across the three domains at each time point. There were moderate to high correlations between the ages 3, 5, and 7 BAS scores ranging from $r = .47$ ($p < .001$) to $r = .59$ ($p < .001$). The BAS was available for 14,374 children at age 3, 14,698 children at age 5, and 13,105 children at age 7.

The FSP (age 5 years). The FSP was completed by teachers when the children were 5 years old. This period corresponded to the end of the 1st year of school for U.K. children in state schools. Three FSP cognitive domains are included in our analysis: (a) communication, language, and literacy; (b) mathematical development; and (c) knowledge and understanding of the world. Teachers rated children's attainment of specific learning goals on a scale from 0 to 9. An overall FSP score was generated using the mean scores across the three cognitive domains. In England, FSP data were collected for the MCS children by the Department for Children, Schools, and Families, whereas in the other parts of the United Kingdom, the FSP data were collected via postal questionnaire sent to individual teachers. These data were then matched to the records of children in the MCS. The differential data collection methods resulted in English children being matched more successfully to the MCS data. Of the 9,084 children remaining in the (English) sample at 5 years, 8,671 were matched. This represents a success rate of around 95%. Response rates were 59% in Scotland, 54% in Wales, and 68% in Northern Ireland. The difference in administration of the questionnaire is captured by a country indicator that is included in all FSP models.

Independent Variables

Child care. Participation in child care during infancy was measured at 9 months with questions asking the mother who looks after the child when the mother is at work (for working mothers) and who else, other than the mother, looks after the child (for nonworking mothers). Mothers were provided with 18 response options and asked how many hours per week the child is looked after by someone else. Two dichotomous variables were created using this information. The first distinguished children who received any type of nonparental care (34%) for a minimum of 9 hr per week from those who did not, with the latter group being categorized as having parental care (66%). This was referred to as "any type of child care versus parental care" in the analysis. The second variable distinguished two types of nonparental care: "center-care versus informal care." This comparison was limited to the 34% children who received a minimum of 9 hr of nonparental child care per week and distinguished children receiving center-based care (nurseries or crèches, 20%) from those receiving informal care (80%) by 9 months. Among children receiving informal care at 9 months, 61% received grandparental care, 21% were cared by nanny or child minders, and 18% were cared by neighbors, relatives, or friends.

In the MCS data, 68% of the children whose mothers were in paid employment for more than 9 hr per week at 9 months were in some form of child care for more than 9 hr per week at 9 months. Among those in any form of child care, 21.6% were in center-based care and 78.4% were in informal care for more than 9 hr per week.

Potential confounding variables. We distinguished between two types of variables that could bias the association between child care and cognitive development: (a) those measured at 9 months and that could have influenced the selection of families into child care and (b) those measured at or after 9 months and that could be influenced by child-care decisions. Variables from the first group were used to create the propensity score in the PSM models. All variables from the first and second groups were included as covariates in both the PSM and OLS regression models.

Variables That Could Influence the Child-Care Decision

The following variables were measured via maternal report during the first assessment at 9 months and include a range of child, family, and maternal characteristics. All these variables were found to be significantly associated with participation in child care during infancy, and most of them also distinguished between infants attending center-based versus informal child care, at the $p < .05$ level (see Table 1).

Child characteristics included the child's ethnicity (Mixed Race, Indian, Pakistani and Bangladeshi, Black, Other vs. White) and maternal reports of the

Table 1

Descriptive Statistics for Child, Family, and Maternal Characteristics, and for Cognitive Outcomes by Child-Care Group

	Parental care by 9 months *n* = 11,074, 66.36% *M* (*SD*)/%	Informal care by 9 months *n* = 4,489, 26.9% *M* (*SD*)/%	Center care by 9 months *n* = 1,124, 6.74% *M* (*SD*)/%
Cognitive outcomes			
Bracken School Readiness age 3	98.38[a] (0.15)	102.80 (0.23)	109.29[b] (0.40)
British Ability Scale (BAS) age 3	98.49[a] (0.15)	103.25 (0.23)	106.36[b] (0.37)
BAS age 5	98.67[a] (0.15)	102.42 (0.23)	107.26[b] (0.39)
Foundation Stage Profiles age 5	98.46[a] (0.15)	102.90 (0.25)	106.98[b] (0.42)
BAS age 7	98.95[a] (0.15)	101.71 (0.24)	106.63[b] (0.42)
Family characteristics			
Annual household income at 9 months (in GBP)	17,841[a] (142)	25,527 (266)	38,461[b] (590)
Annual household income at age 3 (in GBP)	20,056[a] (167)	28,321 (307)	41,473[b] (659)
Annual household income at age 5 (in GBP)	22,584[a] (168)	29,876 (294)	43,774[b] (657)
Annual household income at age 7 (in GBP)	25,342[a] (213)	33,259 (352)	48,990[b] (796)
Living in government-subsidized housing (at 9 months)	23.97%[a]	11.41%	2.16%[b]
Home Observation Measurement of the Environment score (at age 3)	11.92[a] (0.02)	12.27 (0.03)	12.72[b] (0.04)
Parental time investment (at age 3)			
Reading	–0.02[a] (0.01)	0.05 (0.01)	0.00[a] (0.02)
Teaching	–0.04[a] (0.01)	0.07 (0.01)	0.29[b] (0.02)
Maternal characteristics			
Education (at 9 months)			
General Certificate of Secondary Education (GCSE) grades D–G	34.65%[a]	18.06%	4.01%[b]
O level/GCSE grades A–C	37.28%[a]	37.48%	21.69%[b]
A/AS/S Levels	8.97%[a]	11.76%	12.88%
Diploma in higher education	7.57%[a]	11.34%	12.80%
First degree	9.25%[a]	17.15%	37.11%[b]
Higher degree	2.28%[a]	4.22%	11.51%[b]
Employment			
Employed at 9 months	23.90%[a]	84.75%	96.57%[b]
Employed at age 3	35.20%[a]	78.41%	90.83%[b]
Employed at age 5	43.56%[a]	77.55%	86.28%[b]
Employed at age 7	50.56%[a]	78.48%	88.50%[b]
Age at child's birth	27.97[a] (0.06)	28.57 (0.09)	31.58[b] (0.14)
Single at child's birth	10.74%[a]	7.41%	4.11%[b]
Planned pregnancy	51.00%[a]	57.29%	75.76%[b]
Smoked when pregnant	27.54%[a]	20.66%	12.16%[b]
Attended antenatal classes	28.41%[a]	44.75%	60.40%[b]
Ever breastfed	65.68%[a]	67.44%	84.22%[b]
Attachment score (at 9 months)	24.77[a] (0.03)	24.48 (0.04)	23.82[b] (0.07)
Self-esteem score (at 9 months)	18.51[a] (0.03)	18.82 (0.04)	19.02[b] (0.09)
High social class (at 9 months)	19.16%[a]	38.00%	72.52%[b]
Literacy difficulties (at 9 months)	11.45%[a]	8.49%	5.39%[b]
Long-term chronic illness (at 9 months)	22.85%[a]	19.34%	16.27%[b]
Child characteristics			
Female	49.05%	48.98%	47.45%
Ethnic group (% Non-White)	20.53%[a]	11.92%	6.86%[b]
Birth weight (in kg)	3.34[a] (0.01)	3.38 (0.01)	3.44[b] (0.02)
Days/weeks of gestation	277.10/39.59[a] (0.14/0.02)	277.61/39.66 (0.21/0.03)	278.36/39.77 (0.380.05)

Table 1
Continued

	Parental care by 9 months n = 11,074, 66.36% M (SD)/%	Informal care by 9 months n = 4,489, 26.9% M (SD)/%	Center care by 9 months n = 1,124, 6.74% M (SD)/%
Birth by c-section	10.73%[a]	14.17%	16.67%[b]
Birth order	2.13[a] (0.01)	1.65 (0.01)	1.49[b] (0.02)

[a]Significant difference between "parental care" and "any child care," at $p < .05$. [b]Significant difference between "informal child care" and "center-based child care," at $p < .05$.

child's birth weight (in kilograms) and the number of days in gestation. Finally, a binary indicator representing whether the child was delivered by emergency cesarean section was also included.

Family characteristics included the child's rank by age among his or her natural siblings (birth order) and a dummy variable distinguishing families living in government-subsidized housing.

Maternal characteristics and maternal behavior variables included the mother's age at the time of the cohort child's birth, whether the mother was single at the time of the child's birth, and whether the mother had a long-term chronic illness, defined as an illness, disability, or infirmity for a long period of time. In addition, variables indicating the mother's socioeconomic class were included. Mothers who are employed, or who have previously been employed, in higher or lower management or professional jobs were classified as high social class as compared to lower employment classes. Three questions asked the mother if she had difficulties with reading, filling out forms, or calculating change. If she indicated difficulty on any of these tasks, she was characterized as having literacy difficulties. Maternal education was measured using a six-level categorical variable indicating whether the mother's highest educational qualification was a (a) General Certificate of Secondary Education (GCSE) with grades D–G, which is equivalent to 11 years of education; (b) GCSE with grades A–C, which is equivalent to 11 years of education; (c) Advanced Level General Certificate of Secondary Education (A Level), which is equivalent to 13 years of education; (d) diploma in higher education, which is equivalent to 14/15 years of education depending on length of diploma; (e) first degree, which is equivalent to 16 years of education; or (f) a higher degree (e.g., M.A. or Ph.D.), which is equivalent to 17–21 years of education depending on degree type. A number of pregnancy and birth-related variables were also controlled for, including whether the pregnancy was planned, whether the mother smoked during pregnancy after the pregnancy had been confirmed, and whether she attended any antenatal classes during pregnancy. Finally, an indicator of whether the child was ever breastfed was also included.

Variables That Could Be Influenced by the Child-Care Decision

The following variables were also used as covariates in the OLS and PSM models, yet they were not used to generate the propensity scores as they may have been influenced by the child-care decision in infancy. These included a dummy indicator of low income at 9 months, a dummy indicator of maternal employment at 9 months, and household income and maternal employment measured at the same time as the cognitive assessments (i.e., 3, 5, and 7 years). Annual household income was assessed and calculated using the same method at each of the four data collection waves. Respondents were asked to state their net household income from all sources and earnings, after deductions for tax, national insurance, and so forth, from a list of 18 possible income brackets. This information was used to generate a pseudocontinuous measure by taking the midpoints of each income bracket to approximate total household income. There was a high correlation in household income across the four time periods ranging from 0.66 to 0.77 ($p < .01$). A dummy was used for low income at 9 months to facilitate interpretation of child care by low-income interactions (see below for how the variable was derived). Maternal employment was measured at each data collection wave and was coded 1 if the mother said she was in part- or full-time employment and 0 otherwise. Maternal attachment and self-esteem were assessed at 9 months using a short form of the Condon Maternal Attachment Scale (Condon & Corkindale, 1998) and a short form of the Rosenberg Self-Esteem Scale (Rosenberg, 1965), respectively. Higher scores indicated greater attachment and self-esteem.

Additional variables included continuous scores on the Home Observation Measurement of the Environment (HOME) scale (Caldwell & Bradley, 1984), which assessed the quality and quantity of stimulation and support available to a child in their home at age 3 years. A measure of parental time inputs at age 3 included seven items related to how often someone in the home taught the child to learn the alphabet, count, sing, draw, play sports, and how often someone read to the child and took the child to the library. Principal component factor analysis, conducted with the seven items, revealed two independent factors representing reading and teaching, respectively. This information was used to generate two parental investment variables: "Parental time investment: reading" represents the mean of the alphabet, count, sing, draw, sports measures, and "parental time investment: teaching" represents the mean of the reading and library measures.

An indicator of country of residence at the first wave (England, Scotland, Wales, or Northern Ireland) was included in the analysis of the FSP models to account for the differing method of FSP data collection in each country. Finally, the gender of the child and the child's age in months at the time of each cognitive assessment were also included.

Subsequent Child-Care Experiences

The OLS and PSM models (but not the model used to generate the propensity score) also controlled for all subsequent preschool child-care experiences up to ages 3 and 5 years. By age 5 the majority of the sample had started school. The measures of subsequent child-care experiences were retrospective measures collected at the age 5 data collection wave. Mothers were asked at what age the child started informal care (child minder) or center-based care (e.g., nursery school, playgroup, preschool, or day nursery), and for how many hours. This information was used to generate dichotomous variables indicating whether the child had been in any form of child care for at least 9 hr a week by ages 3 (67%) and 5 (95%) years and, for those who have been in child care, whether they received center-based care versus informal care by ages 3 (63% vs. 37%) and 5 (95% vs. 5%) years.

Socioeconomic moderator variables. Maternal education and family income were tested as potential moderators of the relations between child care during infancy and cognitive outcomes. The moderator variables were dichotomized as previous studies identified associations between child care and cognitive development for the lower end of the socioeconomic distribution (e.g., Côté et al., 2007; Geoffroy et al., 2010; Peisner-Feinberg et al., 2001; Waldfogel et al., 2002). Low maternal education indicated that the mother had obtained a GCSE with grades D–G or less (i.e., having the minimum mandatory level of education, approximately equivalent to a high school degree in the United States). Using this definition, approximately 28% of all mothers in the MCS were considered to have low levels of education. Low income indicated that the total annual household income was at or below the U.K. poverty threshold (60% of the median income within the sample adjusted for weighting). With this definition, 28% of the sample had low income. There was a significant overlap between low maternal education and low family income, $F(1, 1893.5) = 1222.96$, $p < .0001$. Whereas 58% of the sample had neither low income nor low education, just 14% had both low income and low education, 14.1% had low income only, and 13.9% had low education only.

Analyses

Comparing child-care conditions. We conducted two main sets of analyses. In the first, we compared the cognitive outcomes of children who received "any type of child care" with those of children who received "parental care" during the first 9 months of life. In the second, we compared the cognitive outcomes of children who received "center-based care" with those of children who received "informal care" during the first 9 months of life. Separate models were used to predict performance on five cognitive tests: the Bracken School Readiness (BSR) and the BAS at age 3, the BAS and the FSP at age 5, and the BAS at age 7 years.

Addressing selection biases. We estimated both OLS and PSM models. First, we estimated OLS models controlling for all covariates listed in the Measures section: gender and age of the child, ethnicity, days of gestation, birth weight, cesarean section, birth order, breastfed, planned pregnancy, antenatal classes, smoked during pregnancy, mother's age, mother single at the birth of the child, mother's education, mother's literacy difficulties, mother's social class, mother has a long-term chronic illness, mother's employment at 9 months, low household income at 9 months, mother's employment and household income at the time of the cognitive assessment, maternal attachment at 9 months, maternal self-esteem at 9 months, living in government-subsidized housing, HOME score at 3 years, parental time investment at 3 years, subsequent child-care use up to age 3 for models of cognitive outcomes at age 3, and up to age 5 for models of cognitive outcomes at ages 5 or 7. The FSP models also included a country dummy.

Second, we estimated PSM models. Propensity scores were estimated for each child to represent the probability of being in any form of child care during infancy and, among those children, their probability of being in center-based versus informal child care. Probit regression models were used, with child-care status as the binary outcome indicator and the variables that preceded the child-care experience as the predictor variables—mother's age at time of child's birth, whether the mother was single at the time of the child's birth, whether the mother had a long-term chronic illness, social class, maternal literacy difficulties, maternal education, planned pregnancy, maternal smoking during pregnancy, attending antenatal classes, breastfeeding, child's ethnicity, birth order, child's birth weight, number of days in gestation, cesarean section delivery, and government-subsidized housing. From the probit models, a predicted score was calculated for each child. This propensity score could theoretically range from 0 to 1 and represented the probability that the child received the child-care "treatment" of interest. We then used a radius matching algorithm to match each child who participated in child care (or in center-based child care for the second set of analyses) with all children who did not, but whose propensity scores lay within a chosen distance (caliper). With radius matching, all untreated individuals within the caliper distance are matched to a treated individual and are weighted equally (i.e., rectangular kernel). As an example, if four untreated individuals are within the specified caliper distance, each is matched to the treated individual and is assigned a weight of .25. With radius matching, an untreated individual may be used as a match for more than one treated individual (akin to matching with replacement). Untreated individuals who are matched to multiple treated individuals consequently have greater overall weight in the analysis compared to those who are matched to fewer treated individuals. Treated individuals for whom no untreated individuals are available within the caliper are excluded from the analysis as this indicates insufficient overlap in the propensity scores to find a suitable "counterfactual." The radius matching procedure often results in more precise estimates of treatment effects than a 1:1 matching algorithm (i.e., nearest neighbor matching), as it allows the use of information from a larger number of untreated participants.

The main advantage of PSM over OLS is that it allows more flexibility in the functional form. Given that the matching estimator simply compares the outcomes of individuals who are observationally similar to one another in terms of covariates (yet who differ in their child-care type), the matching estimator is less parametric in the sense that it imposes little structure in the relation between the covariates, treatment, and the dependent variable. Nevertheless, PSM still does not guarantee causal estimates just as in the case of OLS. In particular, matching estimators only provide causal inference in the cases when (a) selection is on observable child and family characteristics, or (b) balancing the observable covariate distribution also balances the unobservable covariate distribution, in the case where unobservable factors also contribute to selection of child care.

The quality of the matching process was evaluated by comparing the child-care (treated) and matched comparison (untreated) groups on all observed pretreatment characteristics, using the complete-case sample (i.e., before MI). As recommended, we aimed to obtain a difference between the two groups smaller than 5% of a standard deviation (standardized bias; see Caliendo & Kopeinig, 2008). A caliper of 0.1 was found to best balance the treated and untreated groups, for both sets of analyses (any child care vs. parental care, and center-based vs. informal child care).

The OLS and PSM models used to test the association between infancy child care and cognitive outcomes included all covariates (from the first and the second set of variables), to reduce the likelihood of any additional bias not controlled for by the matching and to improve efficiency (Rubin, 1973; Rubin & Thomas, 2000). To test whether child care had a greater or lesser impact on the cognitive development of high-risk children, the OLS and PSM models were estimated three times, first excluding any SES interactions and then including either an interaction between the infancy child-care variable and low maternal education or between the infancy child-care variable and low family income at 9 months.

Note that we do not use sampling weights in the analysis as our matching approach estimates the "average treatment effect on the treated." Therefore, as we do not make inferences about the whole U.K. population, the sample weights are not required.

Addressing biases due to missing data. Deletion of cases or variables due to missing data can lead to substantial bias (Hill et al., 2005). To examine the impact of missing data on our estimations, we estimated our models with and without MI. MI is particularly useful when the data are missing completely at random or when the missing data can be fully described by observed variables in the data set. If unobserved child and family characteristics contributed to generating missing data, then MI may result in some bias of estimates. As Wayman (2003) notes, however, the goal of MI is to capture as much of the mechanism that produced the missing data through observed variables, and these results will likely be better than those produced by listwise deletion (i.e., excluding children with missing data). It should be noted that the estimates produced with MI only have a causal interpretation if selection into child care is based on observable characteristics. Thus, in the event that selection into child care is affected by unobserved child characteristics, even if the mechanism that generated missing data can be fully captured by observed variables, the derived estimates represent correlations, rather than a causal estimates. Although we use multiple observable characteristics of the respondents in the imputation process, it is possible that we are not fully capturing all factors influencing attrition; therefore, our MI results cannot be interpreted as causal effects.

We performed MI by chained equations, using the *mi ice* command in Stata 11.0. Missing values were imputed 50 times using the MCS sampling weight and 78 individual items used to generate the variables included in our models, with the original 18,552 participants. Analyses were run with the resulting 50 "completed" data sets and then pooled with Rubin's combination rules (Rubin, 1987), using *mi estimate.* For the PSM approach, the balance between the data sets was assessed in the nonimputed data, that is, in the original data, by comparing the child-care (treated) and matched comparison (untreated) groups on all observed pretreatment characteristics. If the matching process was successful at balancing the covariate distributions across the two groups, the size of the difference should be smaller than 5% of a standard deviation. Once the appropriate matching algorithm was identified using the nonimputed data, the same one was used for the imputed data. All tables and figures reported in this study are based on the pooled imputed values. Note that MI was only conducted for children who had participated in at least one of the cognitive assessments at ages 3, 5, or 7 years ($n \sim 13{,}000$). An overview of the results using the nonimputed data is presented in the Results section.

Calculation of effect sizes. Cohen's d effect sizes were calculated to describe the magnitude of the effects (in standard deviation units) of child-care use on cognitive outcomes for children of mothers with low and high education (Cohen, 1988). Cohen's d was designed for use where the scores being compared are continuous and normally distributed, a condition that is met in this study. For instance, effect sizes comparing "any type of child care" with "parental care" were calculated as: $d = (X_{\text{child care}} - X_{\text{parental care}})/s$ pooled. Where s pooled $= \sqrt{\{[(n_{\text{child care}} - 1)(s_{\text{child care}}{}^{2}) + (n_{\text{parental care}} - 1)\ (s_{\text{parental care}}{}^{2})]/(n_{\text{child care}} + n_{\text{parental care}} - 2)\}}$. Positive ds indicated higher cognitive scores in the child-care than the parental-care group.

Results

We focus on the results obtained using MI. Table 1 presents the percentages or means and standard deviations for the dependent variables and the covariates by child-care status. Table 1 also presents bivariate associations for each measure indicating statistical differences between the "any child care" and "parental care" groups, and the "center-based care" and "informal care" groups.

There were significant differences across all five dependent variables by child-care status, with children in center-based care typically having higher cognitive scores than those in informal care, and those in informal care having higher scores than children in parental care. In general, children who were in any form of child care during infancy experienced less risk on almost all observed child, family, and maternal characteristics, compared to those in parental care. Similarly, those in center care experienced less risk than those in informal care.

Associations Between Infancy Child-Care Use and Cognitive Outcomes at 3, 5, and 7 Years

The differences in children's cognitive outcomes by child-care status presented in the descriptive statistics suggested that child care, and particularly center-based child care, was associated with higher cognitive scores. However, the observed differences in family, maternal, and child characteristics across child-care groups indicated that social selection into child care may account for this relation. In the next set of analyses, we estimated OLS and PSM models of the association between child care during infancy and cognitive outcomes at 3, 5, and 7 years, and then included an interaction between child care and either low maternal education or low income.

Any type of child care versus parental care. We found no main effects of child care on any of the five cognitive tests, using either the OLS or PSM approaches. In addition, we found no significant interactions between child-care participation and low income using either approach; however, we did identify significant interactions with maternal education for two of the cognitive outcomes. Therefore, we only present the OLS and PSM results of models that included the child care by low maternal education interaction (Table 2). For clarity in the presentation, we omit presenting the coefficients for the control variables and present only the coefficients for child care during infancy, maternal education, and the child care during infancy by low maternal education interaction (full results are available in the online supporting information Appendix S1). The *F* test for each model was statistically significant at the .01 alpha level. All effect sizes are computed from the adjusted means obtained from the PSM models.

The two significant interactions in both the OLS and PSM models indicated that the effect of child care on age 3 test scores (BSR and BAS) varied as a function of maternal education. The association between child care and the BSR at 3 years was larger among children of less educated mothers, $d = 0.34$, 95% CI [0.26, 0.43], than among children of more highly educated mothers where it was not significantly different from zero, $d = 0.04$, 95% CI [0.00, 0.08]. Similarly, the association between child care and the BAS at age 3 years was larger among children of less educated mothers, $d = 0.27$, 95% CI [0.19, 0.36] than among children of more highly educated mothers, $d = 0.10$, 95% CI [0.06, 0.13]. Children in any form of child care did not differ from those in parental care on the ages 5 or 7 years cognitive tests, regardless of maternal education levels.

Center-based child care versus informal care. The OLS and PSM approaches yielded partially different results regarding the effects of center-based versus informal child-care attendance on future cognitive outcomes. Both approaches detected main effects of center care on BSR at 3 years and BAS at 5 years. The effect size for center-based child care versus informal child care, across children attending child care during infancy, was $d = 0.42$, 95% CI [0.34, 0.49] on the BSR at age 3 years and $d = 0.31$, 95% CI [0.24, 0.38] on the BAS at age 5 years.

In addition to the main effect of center-based child care on the BSR at 3 years, there was a Center-Based Child Care × Low Maternal Education interaction in the PSM model only. The association between center-based child care and scores on the BSR at 3 years was larger among children of less educated mothers, $d = 1.15$, 95% CI [0.80, 1.50], than among children of more highly educated mothers, $d = 0.40$, 95% CI [0.32, 0.47].

Finally, there was a significant interaction between center-based child care and maternal education on the FSP at 5 years in the PSM model. The association between center-based child care and the FSP at 5 years was larger among children of less educated mothers, $d = 0.93$, 95% CI [0.53, 1.33] than among children of more highly educated mothers, $d = 0.09$, 95% CI [0.09, 0.17]. Children in center-based child care did not differ from those in informal child care on the BAS at ages 3 or 7, regardless of maternal education level.

With both approaches, we found no significant interactions between child-care participation and low income on any of the five outcomes. However, the PSM approach detected two significant interactions between center-based child-care participation and low maternal education that were not detected with the regression approach. In Table 3, we only present the results of models that included the Center-Based Child Care by Low Maternal Education interaction, with the coefficients for center-based child care during infancy, maternal education, and the Center-Based Child Care During Infancy × Low Maternal Education interaction, omitting coefficients for other control variables (full results are available in online supporting information Appendix S1). The *F* test for each model was statistically significant at the .01 alpha level.

Comparison of results with and without MI. The estimates obtained using listwise deletion (or complete-case analysis) and MI were in the same direction; however, some effects reached statistical significance only with the MI approach. First, there were no differences in the results from the main effect models estimated with the imputed and listwise deletion samples. Second, regarding the interaction models, the same pattern of results was found for the BSR at age 3, BAS at age 5, FSP at age 5, and BAS at age 7 models comparing any type of child care with parental care. However, the interaction effect observed between any form of child care and low maternal education in the BAS age 3 OLS model was not significant when estimated with listwise deletion, yet there was no difference for the PSM models. Third, the same pattern of results was

Table 2

Associations Between the Use of Any Type of Child Care in Infancy, Maternal Education, and Their Interaction, With Cognitive Outcomes at 3, 5, and 7 Years From Ordinary Least Square (OLS) Regression and Propensity Score Matching (PSM) Model

	Age 3 years				Age 5 years				Age 7 years	
	Bracken School Readiness Scores (n = 13,652)		British Ability Scale (BAS; n = 14,374)		BAS (n = 14,793)		Foundation Stage Profile (FSP; n = 11,708)		BAS (n = 13,464)	
	OLS	PSM	OLS	PSM	OLS	PSM	OLS	PSM	OLS	PSM
Any child care in infancy	−0.51 (0.34)	−0.38 (0.38)	−0.73* (0.33)	−0.37 (0.37)	−0.10 (0.35)	0.06 (0.37)	0.05 (0.39)	0.31 (0.40)	−0.50 (0.37)	−0.44 (0.41)
Maternal education[a]										
O level/General Certificate of Secondary Education (GCSE) grades A–C	2.30** (0.33)	2.72** (0.40)	2.15** (0.33)	2.39** (0.39)	2.33** (0.35)	2.27** (0.42)	2.59** (0.38)	2.89** (0.45)	2.68** (0.38)	3.32** (0.45)
A/AS/S Levels	3.50** (0.46)	4.09** (0.54)	2.89** (0.45)	2.89** (0.52)	2.53** (0.49)	2.48** (0.56)	3.91** (0.54)	3.73** (0.58)	4.27** (0.52)	4.81** (0.58)
Diploma in higher education	3.77** (0.48)	4.27** (0.53)	3.56** (0.46)	3.55** (0.51)	3.45** (0.52)	3.32** (0.54)	3.14** (0.55)	3.02** (0.60)	3.28** (0.55)	3.88** (0.59)
First degree	5.17** (0.49)	5.61** (0.56)	4.38** (0.47)	4.42** (0.53)	4.78** (0.52)	4.83** (0.55)	4.76** (0.56)	5.03** (0.58)	5.88** (0.54)	6.90** (0.60)
Higher degree	5.31** (0.69)	5.71** (0.75)	4.14** (0.67)	4.05** (0.75)	4.85** (0.73)	4.86** (0.73)	4.44** (0.80)	4.67** (0.77)	6.31** (0.76)	7.18** (0.80)
Low Maternal Education × Any Child Care	1.32* (0.62)	1.70** (0.64)	1.41* (0.61)	1.46* (0.63)	0.42 (0.65)	0.41 (0.70)	−0.53 (0.71)	−0.42 (0.80)	0.13 (0.69)	0.71 (0.76)

[a] Omitted level: GCSE grades D–G. Coefficients for covariates are reported in the online supporting information Appendix S1. The following variables were included in the propensity score for the PSM models and as covariates in the OLS models: child ethnicity, number of weeks of gestation, birth weight, birth by cesarean section, birth order, ever breastfed, planned pregnancy, attendance at antenatal classes, smoked during pregnancy, mother's age at the child's birth, mother was single at the birth of the child, mother's education level, mother's literacy difficulties, mother's social class, mother had a long-term chronic illness, and local authority housing at 9 months. The following control variables were included as covariates in all the models: any child care before age 3, center child care before age 3, any child care before age 5 (in BAS5, BAS7, FSP7 models only), center child care before age 5 (in BAS5, BAS7, FSP7 only), child age (at time of assessment), child gender, maternal attachment at 9 months, maternal self-esteem at 9 months, Home observation measurement of the environment score at 3 years, the reading and teaching factors of parental time investment at 3 years, household income (at time of assessment), and maternal employment (at time of assessment). The FSP models also include a country dummy.

*$p < .05$. **$p < .01$.

Table 3

Associations Between the Use of Center-Based Child Care in Infancy, Maternal Education, and Their Interaction, With Cognitive Outcomes at 3, 5, and 7 Years From Ordinary Least Square (OLS) Regression and Propensity Score Matching (PSM) Models

	Age 3 years				Age 5 years				Age 7 years	
	Bracken School Readiness Scores (n = 4,260)		British Ability Scale (BAS; n = 4,456)		BAS (n = 4,634)		Foundation Stage Profile (FSP; n = 3,641)		BAS (n = 4,304)	
	OLS	PSM	OLS	PSM	OLS	PSM	OLS	PSM	OLS	PSM
Center-based child care	1.15* (0.53)	1.45** (0.54)	−0.55 (0.49)	−0.24 (0.49)	1.49** (0.53)	1.91** (0.52)	0.63 (0.57)	0.78 (0.51)	0.58 (0.55)	0.71 (0.55)
Maternal education[a]										
O level/General Certificate of Secondary Education (GCSE) grades A–C	1.27 (0.65)	1.37 (0.78)	1.27* (0.61)	1.44 (0.79)	1.72* (0.66)	2.14* (0.86)	2.90** (0.71)	3.13** (0.92)	2.51** (0.69)	2.62** (0.88)
A/AS/S Levels	2.71** (0.82)	3.09** (0.92)	2.21** (0.77)	2.85** (0.92)	2.21** (0.83)	2.68* (1.04)	3.72** (0.90)	4.58** (1.06)	3.89** (0.87)	3.99** (1.04)
Diploma in higher education	2.98** (0.82)	2.37* (0.99)	2.43** (0.77)	2.55** (0.94)	2.17* (0.85)	2.66** (1.00)	2.55** (0.91)	2.52* (1.10)	2.10* (0.88)	2.26* (1.03)
First degree	4.02** (0.83)	4.36** (0.92)	2.83** (0.78)	3.69** (0.92)	3.43** (0.85)	3.87** (1.00)	4.33** (0.92)	4.63** (1.06)	4.84** (0.88)	4.68** (1.03)
Higher degree	3.67** (1.08)	3.90** (1.17)	3.32** (1.01)	4.42** (1.18)	3.44** (1.09)	3.87** (1.18)	4.45** (1.17)	4.81** (1.23)	6.97** (1.13)	6.53** (1.23)
Low Maternal Education × Any Child Care	4.09 (2.14)	3.94* (1.80)	1.67 (2.02)	1.77 (1.67)	2.32 (2.26)	2.57 (2.37)	4.57 (2.55)	4.50* (1.99)	1.90 (2.40)	1.85 (2.71)

[a] Omitted level: GCSE grades D–G. Coefficients for covariates are reported in the online supporting information Appendix S1. The following variables were included in the propensity score for the PSM models and as covariates in the OLS models: child ethnicity, number of weeks of gestation, birth weight, birth by cesarean section, birth order, ever breastfed, planned pregnancy, attendance at antenatal classes, smoked during pregnancy, mother's age at the child's birth, mother was single at the birth of the child, mother's education level, mother's literacy difficulties, mother's social class, mother had a long-term chronic illness, and local authority housing at 9 months. The following control variables were included as covariates in all the models: any child care before age 3, center child care before age 3, any child care before age 5 (in BAS5, BAS7, FSP7 models only), center child care before age 5 (in BAS5, BAS7, FSP7 only), child age (at time of assessment), child gender, maternal attachment at 9 months, maternal self-esteem at 9 months, Home observation measurement of the environment score at 3 years, the reading and teaching factors of parental time investment at 3 years, household income (at time of assessment), and maternal employment (at time of assessment). The FSP models also include a country dummy.

$*p < .05$. $**p < .01$.

found regarding BAS at ages 3, 5, and 7 in the center-based versus informal-care models that included the interaction between center-based child care and low maternal education. However, the interaction effects observed for the BSR age 3 and FSP age 5 PSM models were not significant when estimated with listwise deletion, yet there was no difference in the OLS models. For almost all estimated models, both the coefficients and the standard errors were larger in the listwise deletion results compared to the imputed results reported in Tables 2 and 3. Regarding the three models where differences were observed between listwise deletion and MI, in all cases, both the coefficients and standard errors were lower in the models estimated with the imputed data. The listwise deletion results are available from the authors upon request.

Discussion

The aim of this study was to investigate whether nonparental child care received in infancy was associated with cognitive outcomes at 3, 5, and 7 years among a large (*n* =~13,000) sample of children born in the United Kingdom between 2000 and 2002. We found that child-care use at 9 months was associated with better cognitive development for certain outcomes, yet the effects depended on the level of maternal education. The significant effects of any type of child care were small to moderate and were limited to children of mothers with low levels of education and to the preschool years (i.e., age 3 years). We found that center-based care, compared to informal care at 9 months, was associated with better cognitive development for certain outcomes, and again the effects varied as a function of maternal education. The significant effects were moderate overall and large for children of mothers with low education. The effects for center-based child care extended until school entry (5 years), but not beyond (age 7 years). The results inform the ongoing debates about the use of child-care services in infancy by indicating that services received at this early stage of life did not carry risk, but rather made no difference or contributed positively to children's cognitive development, in the case of children of mothers with low levels of education.

Although we found a short-term advantage of participating in any type of child care for children of mothers with low education, our results suggest that center-based child care increases further the likelihood of beneficial effects for these children. The characteristics of center-based settings may make them more likely than those of informal-care settings to foster cognitive abilities. Informal care is usually family based and aimed at providing a "home-like" setting that involves mostly free play. Infants receiving informal care are exposed to an environment with less focus on structured activities than children in center-based care. Center care is usually provided to groups of children of similar ages and there is evidence that age segregation may promote higher quality care and education (NICHD ECCRN, 2004). Indeed, in groups of children of similar ages, the type of care and activities are more easily targeted to children's developmental needs. For instance, in mixed-age settings, younger children (i.e., toddlers) were shown to receive less sensitive and supportive care than older ones (i.e., preschoolers; Kryzer, Kovan, Phillips, Domagall, & Gunnar, 2007). Thus, children who experience center care have more exposure to an environment designed as an educational setting and have access to a greater number and variety of toys and materials for children (Kisker & Maynard, 1991). Our results suggest that the characteristics of center-based care provide an advantage even when child care is initiated during infancy. Our results are consistent with previous studies reporting a positive effect for center-based services in infancy (e.g., Gregg et al., 2005; Hansen & Hawkes, 2009; Sylva et al., 2011), but contrary to some of these studies (e.g., Gregg et al., 2005; Sylva et al., 2011), we did not find a negative effect of informal care and even found an overall positive effect of child care for children of mothers with low education. However, these studies did not examine maternal education as a moderator of informal child care.

Another objective of this study was to test the extent to which there was a differential effect of child care, and the type of child care, according to socioeconomic risk. We examined both low maternal education and low income as potential moderators to shed light on previously divergent results in studies using one or the other moderator. We found differential impacts of child care and the type of child care according to maternal education. All four significant interactions revealed the same pattern, whereby children of mothers with low education particularly benefited from child-care services, particularly from center-based rather than informal child care, whereas children of mothers with higher levels of education benefited modestly, moderately, or not at all. Indeed, we found that children of mothers with low levels of education receiving any type of child care had higher scores on the BSR and on the BAS at 3 years compared to children of mothers with low education receiving parental care. The effect sizes were moderate. Children of mothers with low education in center-based child care had higher scores on the BSR at 3 years and on the FSP at 5 years compared to children of mothers with low education receiving informal child care, and the effects sizes were large. Although the effect sizes were smaller (i.e., small to moderate), center-based care was also found to be associated with higher scores on some cognitive tests at 3 and 5 years among children of mothers who did not have low levels of education. These results are consistent with a previous study of Canadian children (Geoffroy et al., 2010) and suggest that child care, particularly center-based child care, may attenuate the cognitive disadvantage associated with growing up in a family where the mother has low education.

Despite the significant overlap between low maternal education and low family income in this study, $\chi^2(1) = .0014$, $p < .0001$, differential effects of child care were not found when low income was used as the moderator. This suggests that child care may serve to partially compensate for a lack of stimulation (low maternal education) rather than material deprivation (low income). These results underscore the importance of distinguishing the moderating role of low maternal education and low income in child-care studies and suggest that children of mothers with low levels of education may benefit from child-care services more than children of low-income families.

Accounting for Methodological Biases

Given the previously documented impact of social selection and attrition in longitudinal child-care studies (Hill et al., 2005), we carefully addressed these biases using PSM and MI. PSM and MI yielded significant findings that were undetected using the regression and complete-case approaches. In addition, we controlled for maternal employment, a variable often omitted in child-care research.

There were strong selection effects in the use of child care. Children exposed to higher levels of family and maternal risk characteristics were much more likely to receive parental-care or informal child-care services. The magnitudes of the effects obtained with OLS regression analyses were largely consistent with those obtained with PSM analyses. However, in two instances, we detected significant interactions between maternal education and center-based child care using PSM that we did not detect using OLS. This suggests that the assumptions of OLS may not be tenable in the presence of strong selection effects and may conceal group differences.

We also observed differences when we compared the results from listwise deletion (i.e., complete-case approach) with those obtained with MI. Some of the Maternal Education × Child-Care interactions detected with MI did not reach significance using the complete-case approach. The MI approach brought more of the disadvantaged children (who are typically disproportionately affected by attrition) into the analyses and therefore provided adequate power for testing SES interactions, particularly for the center-based versus informal-care analyses. This pattern of results is consistent with Hill et al. (2005) who reported that MI reduced the negative effect of maternal employment (compared to the nonimputed results).

Strengths and Limitations

This study has a number of considerable strengths. First, it is based on one of the largest samples available to examine the association between child care and cognitive outcomes. As noted by Berger et al. (2008), sample size is an important limitation of child-care studies. The MCS is more than 10 times larger than the state-of-the-art studies and thus allows for more precise estimations of the associations between child care and cognitive development. The second strength concerns the extensiveness of the cognitive assessments: (a) variety of cognitive domains: school readiness, vocabulary, communication, mathematical skills, knowledge of the world, reading, number skills, and pattern construction; (b) developmental scope: 3, 5, and 7 years; and (c) source of assessment: formal testing and teacher ratings. A third strength is found in our use of PSM and OLS to control for selection into child care based on observed characteristics. A fourth strength is the fact that our results allow us to distinguish the role of child care from that of maternal employment as we control for maternal employment. Indeed, we found that child care still had an impact on cognitive development while controlling for maternal employment. Finally, our use of MI reduced the biases associated with differential attrition that can be accounted for by observed characteristics.

There are also limitations to our study. First, we could not account for the quality of child care. As with other large-scale studies designed to examine the role of numerous factors in child development, detailed measures of child-care quality and family processes were not available. Previous studies suggest that child-care type is associated with child-care quality, although important variations in quality within each child-care type were noted (Japel, Tremblay, & Côté, 2005). Second, MI only corrects the bias due to differential attrition if the missing values can be fully described by observed variables, as discussed above. Thus, if unobserved characteristics affect the mechanism generating missing values, then some bias is still likely to remain. Third, despite our sample being one of the largest available to study child care and child development, our capacity to study interactions between maternal education and the use of center-based child care in infancy was limited (due to low numbers of mothers with low levels of education using center care). Caution is required in interpreting these results. However, the fact that we detected two significant interactions suggests that this is a meaningful association and that it should be the subject of further investigations. Finally, as mentioned previously, PSM is a method for dealing with selection on observable family and child characteristics in a manner that is more flexible than OLS in terms of the functional form; it generally does not control for selection on unobservable (or unmeasured) characteristics. It is possible that there are additional factors that influence both the decision to use child care and children's cognitive development that were not observed, such as parental intelligence, motivation, personality, and educational beliefs. To the extent that such variables were not accounted for by PSM, the results may still be affected by selection bias, thus the estimates do not have a causal interpretation. To minimize the likelihood of unaccounted selection into child care, many additional covariates were included in this analysis. Other methods such as instrumental variables, fixed-effects, or within-family models, which can control for selection on both observable and unobservable characteristics, could not be used in this study. Using a sibling comparison design, Jaffee et al. (2011) found that whereas between-family comparisons showed a positive impact of early child care on childhood and adolescence achievement scores, within-family comparisons failed to detect these effects. This suggests the existence of unmeasured characteristics that may not be adequately controlled for in an OLS model. A study using a sibling comparison design with both OLS and PSM is needed to determine whether the results with PSM converge with those of the sibling comparison.

In sum, we did not detect detrimental effects on cognitive development of early nonparental child care, or of center-based versus informal child care, as provided in the United Kingdom between 2000 and 2003. To the contrary, we showed improved short-term outcomes for children of mothers with low levels of education who received nonparental care during infancy. Center-based care had positive effects over informal care, and these benefits lasted longer and extended to children of educated mothers. Our findings suggest that social programs aimed at providing child-care services to vulnerable populations should target low maternal education rather than low family income. Experimental studies comparing the effect of early child care for children of mothers with low-education versus those of low-income families would prove informative regarding the feasibility of this approach and the potential additional benefits of targeting services based on the strongest predictor of cognitive development–maternal education.

References

Andersson, B. E. (1989). Effects of public day-care—A longitudinal-study. *Child Development*, *60*, 857–866. doi:10.1111/j.1467-8624.1989.tb03518.x

Baydar, N., & Brooks-Gunn, J. (1991). Effects of maternal employment and child-care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from the Children of the National Longitudinal Survey of Youth. *Developmental Psychology*, *27*, 932–945.

Belsky, J. (2001). Emanuel Miller Lecture: Developmental risks (still) associated with early child care. *Journal of Child Psychology and Psychiatry*, *42*, 845–859. doi:10.1111/1469-7610.00782

Berger, L., Brooks-Gunn, J., Paxson, C., & Waldfogel, J. (2008). First-year maternal employment and child outcomes: Differences across racial and ethnic groups. *Children and Youth Services Review*, *30*, 365–387. doi:10.1016/j.childyouth.2007.10.010

Berger, L. M., & Waldfogel, J. (2004). Maternity leave and the employment of new mothers in the United States. *Journal of Population Economics*, *17*, 331–349. doi:10.1007/s00148-003-0159-9

Bowlby, J. (1951). *Maternal care and mental health*. Geneva, Switzerland: World Health Organization.

Bracken, B. A. (1984). *Bracken Basic Concept Scale*. Chicago: Psychological Corporation.

Brooks-Gunn, J., Han, W.-J., & Waldfogel, J. (2002). Maternal employment and child cognitive outcomes in the first three years of life: The NICHD Study of Early Child Care. *Child Development*, *73*, 1052–1072. doi:10.1111/1467-8624.00457

Burchinal, M. R., & Clarke-Stewart, K. A. (2007). Maternal employment and child cognitive outcomes: The importance of analytic approach. *Developmental Psychology*, *43*, 1140–1155. doi:10.1037/0012-1649.43.5.1140

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science, 4*, 149–165. doi:10.1207/S1532480XADS0403_4

Burchinal, M., Roberts, J. E., Zeisel, S. A., Hennon, E. A., & Hooper, S. (2006). Social risk and protective child, parenting, and child care factors in early elementary school years. *Parenting-Science and Practice, 6*, 79–113. doi:10.1207/s15327922par0601_4

Caldwell, B., & Bradley, R. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas at Little Rock.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*, 31–72. doi:10.1111/j.1467-6419.2007.00527.x

Caughy, M. O. B., DiPietro, J. A., & Strobino, D. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development, 65*, 457–471. doi:10.1111/j.1467-8624.1994.tb00763.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.

Condon, J. T., & Corkindale, C. J. (1998). The assessment of parent-to-infant attachment: Development of a self-report questionnaire instrument. *Journal of Reproductive and Infant Psychology, 16*, 57–76. doi:10.1080/02646839808404558

Côté, S. M., Boivin, M., Nagin, D. S., Japel, C., Xu, Q., Zoccolillo, M., et al. (2007). The role of maternal education and non-maternal care services in the prevention of children's physical aggression. *Archives of General Psychiatry, 64*, 1305–1312. doi:10.1001/archpsyc.64.11.1305

Côté, S. M., Borge, A. I., Geoffroy, M.-C., Rutter, M., & Tremblay, R. E. (2008). Nonmaternal care in infancy and emotional/behavioral difficulties at 4 years old: Moderation by family risk characteristics. *Developmental Psychology, 44*, 155–168. doi:10.1037/0012-1649.44.1.155

Crosby, D. A., & Hawkes, D. D. (2007). Cross-national research using contemporary birth cohort studies: A look at early maternal employment in the UK and USA. *International Journal of Social Research Methodology, 10*, 379–404. doi:10.1080/13645570701677151

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development, 80*, 1329–1349. doi:10.1111/j.1467-8624.2009.01336.x

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*, 151–161. doi:10.1162/003465302317331982

Dex, S., & Joshi, H. (Eds.) (2005). *Children of the 21st century: From birth to nine months*. Bristol, England: Policy Press.

Early, D. M., & Burchinal, M. R. (2001). Early childhood care: Relations with family characteristics and preferred care characteristics. *Early Childhood Research Quarterly, 16*, 475–497. doi:10.1016/S0885-2006(01)00120-X

Elliott, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales second edition (BAS II): Administration and scoring manual*. London: NFER-Nelson.

Elliott, C. D., Smith, P., & McCulloch, K. (1997). *British Ability Scales second edition (BAS II): Technical manual*. London: NFER-Nelson.

Geoffroy, M.-C., Côté, S. M., Borge, A., Larouche, F., Séguin, J. R., & Rutter, M. (2007). Association between nonmaternal care in the first year of life and children's receptive language skills prior to school entry: The moderating role of the socioeconomic status. *Journal of Child Psychology and Psychiatry, 48*, 490–497. doi:10.1111/j.1469-7610.2006.01704.x

Geoffroy, M. C., Côté, S. M., Giguere, C. E., Dionne, G., Zelazo, P. D., Tremblay, R. E., et al. (2010). Closing the gap in academic readiness and achievement: The role of early childcare. *Journal of Child Psychology and Psychiatry, 51*, 1359–1367. doi:10.1111/j.1469-7610.2010.02316.x

Gregg, P., Washbrook, E., Propper, C., & Burgess, S. (2005). The effects of a mother's return to work decision on child development in the UK. *Economic Journal, 115*, F48–F80. doi:10.1111/j.0013-0133.2005.00972.x

Han, W.-J., Waldfogel, J., & Brooks-Gunn, J. (2001). The effects of early maternal employment on later cognitive and behavioral outcomes. *Journal of Marriage and Family, 63*, 336–354.

Hansen, K., & Hawkes, D. (2009). Early childcare and child development. *Journal of Social Policy, 38*, 211–239. doi:10.1017/S004727940800281X

Harvey, E. (1999). Short-term and long-term effects of early parental employment on children of the National Longitudinal Survey of Youth. *Developmental Psychology, 35*, 445–459.

Hill, J. L., Waldfogel, J., Brooks-Gunn, J., & Han, W. J. (2005). Maternal employment and child development: A fresh look using newer methods. *Developmental Psychology, 41*, 833–850. doi:10.1037/0012-1649.41.6.833

Jaffee, S. R., Van Hulle, C., & Rodgers, J. L. (2011). Effects of nonmaternal care in the first 3 years on children's academic skills and behavioral functioning in childhood and early adolescence: A sibling comparison study. *Child Development, 82*, 1076–1091. doi:10.1111/j.1467-8624.2011.01611.x

Japel, C., Tremblay, R. E., & Côté, S. (2005). La qualité des services à la petite enfance: résultats de l'Étude longitudinale sur le Développement des Enfants du Québec (ÉLDEQ) [The quality of child care services during early childhood: Results from the Quebec Longitudinal Study on Child Development (QLSCD)]. *Éducation et Francophonie, 23*, 7–27.

Kisker, E. E., & Maynard, R. (1991). Quality, cost, and parental choice of child care. In D. M. Blau (Ed.), *The economics of child care* (pp. 127–143). New York: Russell Sage Foundation.

Kryzer, E. M., Kovan, N., Phillips, D. A., Domagall, L. A., & Gunnar, M. R. (2007). Toddlers' and preschoolers' experience in family day care: Age differences and behavioral correlates. *Early Childhood Research Quarterly, 22*, 451–466. doi:10.1016/j.ecresq.2007.08.004

Lerner, J. V., & Galambos, N. L. (1986). Temperament and maternal employment. *New Directions for Child Development, 31*, 75–88.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66. doi:10.1016/j.econedurev.2005.11.005

NICHD, & Duncan, G. J. (2003). Modeling the impact of child care quality on children's preschool cognitive develepement. *Child Development*, *74*, 1454–1475.

NICHD ECCRN. (1997). Familial factors associated with the characteristics of nonmaternal care for infants. *Journal of Marriage & the Family*, *59*, 389–408.

NICHD ECCRN. (2000a). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science*, *4*, 116–135.

NICHD ECCRN. (2000b). The relation of child care to cognitive and language development. *Child Development*, *71*, 960–980. doi:10.1111/1467-8624.00202

NICHD ECCRN. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*, 133–164.

NICHD ECCRN. (2004). Type of child care and children's development at 54 months. *Early Childhood Research Quarterly*, *19*, 203–230.

NICHD ECCRN. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist*, *61*, 99–116. doi:10.1037/0003-066X.61.2.99

Organisation for Economic Co-operation and Development. (2011). *OECD family database*. Retrieved March 9, 2012, from www.oecd.org/social/family/database.

Panter, J. E. (2000). Validity of the Bracken Basic Concept Scale-Revised for predicting performance on the Metropolitan Readiness Test-Sixth Edition. *Journal of Psychoeducational Assessment*, *18*, 104–110.

Park, K. J., & Honig, A. S. (1991). Infant child care patterns and later teacher ratings of preschool behaviors. *Early Child Development and Care*, *68*, 89–96. doi:10.1080/0300443910680108

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*, 1534–1553.

Plewis, I. (2004). *Millennium Cohort Study: Technical report on sampling*. London: Institute of Education, University of London.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rubin, D. B. (1973). The use of matching and regression adjustments to remove bias in observational studies. *Biometrics*, *29*, 185–203.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*, 573–585.

Ruhm, C. J. (2004). Parental employment and child cognitive development. *Journal of Human Resources*, *39*, 155–192. doi:10.2307/3559009

Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence—The influence of social and family risk-factors. *Child Development*, *64*, 80–97. doi:10.1111/j.1467-8624.1993.tb02896.x

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Educational Research Association.

Singer, J. D., Fuller, B., Keiley, M. K., & Wolf, A. (1998). Early child-care selection: Variation by geographic location, maternal characteristics, and family structure. *Developmental Psychology*, *34*, 1129–1144.

Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L. E., & FCCC-Team. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study. *British Journal of Developmental Psychology*, *29*, 18–45. doi:10.1348/026151010x533229

UNICEF. (2008). *The child care transition: A league table of early childhood education and care in economically advanced countries*. Innocenti Report Card 8. Retrived August 1, 2011, from http://www.unicef-irc.org/publications/507

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, *75*, 296–312.

Waldfogel, J. (2001). International policies toward parental leave and child care. *Future of Children*, *11*, 99–111.

Waldfogel, J., Han, W. J., & Brooks-Gunn, J. (2002). The effects of early maternal employment on child cognitive development. *Demography*, *39*, 369–392.

Wayman, J. (2003). *Multiple imputation for missing data: What is it and how can I use it?* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2011). Head Start and urban children's school readiness: A birth cohort study in 18 cities. *Developmental Psychology*, *47*, 134–152. doi:10.1037/a0020784

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Coefficients for Models of the Associations Between the Use of Child Care in Infancy, Maternal Education, and Their Interaction, With Cognitive Outcomes (for Ordinary Least Squares Regressions and Propensity Score Matching Models).

Child Development, July/August 2013, Volume 84, Number 4, Pages 1209–1225

Genetic Moderation of Early Child-Care Effects on Social Functioning Across Childhood: A Developmental Analysis

Jay Belsky
University of California, Davis and King Abdulaziz University, and Birkbeck University of London

Michael Pluess
Institute of Psychiatry, King's College of London

Data from 508 Caucasian children in the NICHD Study of Early Child Care and Youth Development shows that the DRD4 (but not 5-HTTLPR) polymorphism moderates the effect of child-care quality (but not quantity or type) on caregiver-reported externalizing problems at 54 months and in kindergarten and teacher-reported social skills at kindergarten and first grade—but not thereafter. Only children carrying the 7-repeat allele proved susceptible to quality-of-care effects. The behavior–problem interactions proved more consistent with diathesis-stress than differential-susceptibility thinking, whereas the reverse was true of the social-skills' results. Finally, the discerned Gene × Environment interactions did not account for previously reported parallel ones involving difficult temperament in infancy.

Debate has long characterized discussion of the effects of child care on children's development (Fox & Fein, 1990; Karen, 1998), although this is clearly more true regarding adverse effects of lots of time spent in child care (Belsky, 1986, 1988, 1990; Clarke-Stewart, 1989; Phillips, McCartney, Scarr, & Howes, 1987; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010), perhaps particularly centers (Belsky, Vandell, Burchinal, Clarke-Stewart, McCartney, & Owen, 2007), than with respect to quality of care. After all, it is not only widely believed that quality of child care is an important determinant of children's functioning, but there is long-standing evidence consistent with this claim (e.g., Howes, 1988; Peisner-Feinberg & Burchinal, 1997; Vandell, Henderson, & Wilson, 1988).

Important to appreciate, however, is that inconsistency exists within the research literature even with respect to the effects of child-care quality. Perhaps most notable in this regard is the general failure of careful measurements of how attentive, responsive, and stimulating caregivers were to children enrolled in the large-scale NICHD Study of Early Child Care and Youth Development (SECCYD) to predict children's social adjustment after age 3 and before age 15, even as it consistently predicted children's cognitive-linguistic functioning and academic achievement during this developmental period (Belsky et al., 2007; NICHD ECCRN, 2003; Vandell et al., 2010).

For the past decade or so, developmentalists studying the effects of diverse environmental experiences and exposures have become ever more aware that individuals may differ in whether and how they are affected by their developmental experiences and, most importantly for the research reported herein, that this may be a function of genetics (Caspi & Moffitt, 2006). Although some have questioned the replicability of particular Gene × Environment (G×E) interactions that have appeared in the published literature (Duncan & Keller, 2011; Risch et al., 2009), it would certainly seem mistaken to throw the genetic-moderation-of-environmental-influences baby out with the bathwater (Caspi, Hariri, Holmes, Uher, & Moffitt, 2010; Karg,

---

The research described herein was supported by a cooperative agreement with the National Institute of Child Health and Human Development (NICHD; U10-HD25420). This article was the result of a collaboration of the two named authors using data collected under the direction of the NICHD network authors that has been placed in the public domain. The NICHD network of authors merits our appreciation for ensuring that these data were gathered and made available to all network authors and others who received permission to conduct scientific studies using them. The NICHD network authors, however, have no responsibility for how we have analyzed the data, the results we report, and the conclusions we draw. Special thanks is extended to Glenn Roisman for his herculean efforts in overseeing the genotyping of the sample.

Correspondence concerning this article should be addressed to Jay Belsky, Human and Community Development, University of California, Davis, One Shields Avenue, Hart Hall, Davis, CA 95616. Electronic mail may be sent to jbelsky@ucdavis.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0008
DOI: 10.1111/cdev.12058

Burmeister, Shedden, & Sen, 2011; Rutter, Thapar, & Pickles, 2009; Uher & McGuffin, 2010). Indeed, here we take advantage of data from the NICHD SECCYD to examine for the first time whether two particular genetic factors might moderate effects of quality, quantity, and type of child care on children's teacher-rated externalizing behavior problems and social skills. In so doing, we are positioned to shed some light on at least one reason why child-care effects have been subject to so much debate. After all, if at least some child-care effects prove to be genetically moderated, varying across studies due to the genetic makeup of samples, this could account for some of the inconsistency in the literature. To be appreciated, however, is that because the sample studied here is limited in size, is not epidemiological in character, and we do not have a second sample on which to replicate any detected G×E effects, the research presented herein represents a proof-of-concept effort. While it can shed light on whether child-care effects could vary as a function of children's genotype, it is not positioned to provide compelling evidence that they certainly do. Additional research will be required before such could be claimed.

The Moderation of Child-Care Effects

As it turns out, there is evidence that some child-care effects vary as a function of children's characteristics of individuality. Pluess and Belsky (Belsky & Pluess, 2012; Pluess & Belsky, 2009, 2010) found, upon further analyzing NICHD SECCYD data, that early temperament played a moderational role with respect to effects of quality of care, and it did so in ways consistent with some other work evaluating this possibility (Crockenberg, 2003; Phillips, Fox, & Gunnar, 2011). More specifically, children who were rated as more difficult to care for in their first 6 months of life, including being more negatively emotional, proved to be more affected by quality—but not quantity or type—of care than did other children, at least with respect to externalizing problems and social skills. Just as intriguing was the fact that such children were affected in a for-better-*and*-for-worse manner (Belsky, Bakermans-Kranenburg, et al., 2007). That is, children with histories of difficult temperament in infancy scored higher on behavior problems than all other children across the preschool, middle-childhood, and adolescent years if they had experienced low-quality care, yet lower on behavior problems than others if they had experienced high-quality care (Belsky & Pluess, 2012; Pluess & Belsky, 2009, 2010).

Models of Environmental Action

Such moderational effects of temperament vis-à-vis child-care experiences were particularly interesting in that they proved more consistent with the differential-susceptibility hypothesis (Belsky, Bakermans-Kranenburg, et al., 2007; Belsky & Pluess, 2009; Ellis, Boyce, Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2011) than the prevailing diathesis-stress model of environmental action (Zuckerman, 1999). The long-standing and highly influential diathesis-stress framework exclusively addresses the issue of who is most susceptible to the *negative* effects of *adverse* environmental experiences and exposures (e.g., poverty, harsh parenting), essentially presuming that some are more *vulnerable* than others as a result of a personal characteristic (e.g., difficult temperament, risk allele). It is this conceptual framework that informed the first G×E interaction research examining the effects of maltreatment on antisocial behavior (Caspi et al., 2002) and of life event stress on depression (Caspi et al., 2003).

Notably, the diathesis-stress paradigm does not in any way address differential response to positive, supportive, or enriching environmental experiences and exposures (e.g., high-quality child care, sensitive parenting), and this is where it differs from the differential-susceptibility perspective. Rather than regarding some individuals as exclusively more vulnerable than others to adversity due to some endogenous factor, the differential-susceptibility model of environmental action presumes that those most likely to be negatively affected by some adverse environmental condition are also most likely to *benefit* from a *supportive* one. Thus, whereas diathesis-stress thinking regards some as more vulnerable to adversity, differential-susceptibility thinking stipulates that those most susceptible to negative environmental influences are simultaneously most susceptible to positive ones as well. That is, they are not so much vulnerable as malleable or developmentally plastic—for-better-*and*-for-worse. Intriguingly, a good deal of G×E interaction work, the focus of the current report, provides evidence consistent with differential-susceptibility theorizing (Belsky et al., 2009; Belsky & Pluess, 2009; Ellis et al., 2011).

G×E Interaction

In the current report, we extend examination of the potential moderating effect of child attributes with respect to the potential influence of day care (i.e., quality, quantity, and type) on teacher-rated social adjustment by focusing on children's genetic makeup. In particular, in the G×E analyses reported herein, we explore the prospect that the serotonin transporter gene, 5-HTTLPR, or the DRD4 dopamine-receptor gene moderate the effect of child-care experience on teacher-rated externalizing problem behavior and social skills, perhaps accounting for why quality of care in particular did not predict—as a main effect—social adjustment from just before school entry to the last year of elementary school in the NICHD SECCYD (Belsky et al., 2007; Belsky, Bakermans-Kranenburg, et al., 2007). We restrict our analyses to these two polymorphisms, despite the fact that a range of VNTRs and SNPs have been genotyped on the SECCYD cohort, because they have proven most often in G×E research to function as "plasticity genes" in line with differential-susceptibility rather than just "risk alleles" consistent with diathesis-stress models of environmental action (Belsky & Pluess, 2009).

There are several additional reasons to focus on the DRD4 polymorphism in this study of the effects of quality, quantity, and type of care on externalizing problem behavior and social skills. First, DRD4 has been linked with aggression (and antisocial behavior) in children and adolescents (Beaver et al., 2007; Oades et al., 2008; Schmidt, Fox, Rubin, Hu, & Hamer, 2002), a core component of externalizing problems that is one of the two developmental outcomes that are the focus of this inquiry and that has been the source of so much controversy with respect to day-care effects (Crockenberg, 2003; Fox, 1990; Langlois & Liben, 2003; Maccoby & Lewis, 2003). Second, the DRD4 polymorphism has been related to negative emotionality in infancy (Holmboe, Nemoda, Fearon, Sasvari-Szekely, & Johnson, 2011), thereby raising the prospect that the repeatedly chronicled and aforementioned Temperament × Child Care quality interaction discerned by Pluess and Belsky (Belsky & Pluess, 2012; Pluess & Belsky, 2009, 2010) could be the result of a G×E interaction involving DRD4; we evaluate this very possibility in the final analysis presented in this report.

A third and perhaps more important reason for focusing on the DRD4 polymorphism in this inquiry is that Dutch investigators have found that it moderates the effect of quality of parenting on externalizing behavior problems in both observational research (Bakermans-Kranenburg & van IJzendoorn, 2006) and intervention work (Bakermans-Kranenburg, van IJzendoorn, Pijlman, Mesman, & Juffer, 2008). Specifically, children carrying the 7-repeat allele proved more susceptible to parenting effects than did children not carrying this version of the gene, and they did so in a manner consistent with differential susceptibility. Because the measurements of quality of child care in the NICHD study tap much the same caregiving behavior of child-care providers as did the measures of parenting examined in these Dutch studies, it certainly is conceivable that this same polymorphism might moderate the effects of child-care quality—and perhaps quantity and type of care as well—on externalizing behavior problems and social skills. This possibility would seem further buttressed by the fact that a 10-study meta-analysis revealed that variation in genes related to dopamine signaling in the brain influence children's sensitivity to both sensitive or responsive *and* harsh or unresponsive parenting (Bakermans-Kranenburg & van IJzendoorn, 2011).

Reasons for focusing on the serotonin transporter gene, 5-HTTPLR, as a possible moderator of child-care effects in this report are much the same as those highlighted in the case of DRD4. It, too, has been specifically implicated in moderating effects of parenting on child development (e.g., Barry, Kochanska, & Philibert, 2008) and has been found to be related to negative emotionality (Auerbach et al., 1999), even moderating the effect of prenatal stress on negative emotionality in infancy (Pluess et al., 2011). In fact, a recent meta-analysis of G×E studies involving children under 18 years of age reveals 5-HTTLPR to operate as a "plasticity gene"—with carriers of short alleles disproportionately benefiting from supportive environmental conditions while also being most adversely affected by negative contextual conditions more than others (van IJzendoorn, Belsky, & Bakermans-Kranenburg, in press).

Current Study

In sum, in this work we address, for the first time, whether effects of day care are genetically moderated, by focusing on two particular polymorphisms, DRD4 and 5-HTTLPR, and social adjustment as measured by repeated teacher reports of externalizing behavior problems and social skills across the middle-childhood years. It should be noted that this study involves a smaller sample than used in most analyses of NICHD SECCYD data due to the fact that not all children provided DNA. Although this makes precise comparison with other NICHD SECCYD reports difficult, it is not unusual for parents to turn down the opportunity for genetic data to be collected on their children.

We further seek to determine whether any detected G×E effects prove consistent with diathesis-stress or differential-susceptibility theorizing. In addition, we endeavor to determine whether conclusions that only quantity and type of child care, not quality of care, predict social adjustment, at least during the period from just before or at time of school entry to the end of the elementary school years, need to be modified as a result of the detection of G×E interaction involving quality of care. In so doing, we extend prior work showing that at least one child attribute, early difficult temperament, moderates the effect of child-care quality on externalizing problems and social skills. Indeed, we additionally seek to determine, to our knowledge for the first time, whether a previously documented moderational effect of a feature of temperament is independent of genetic moderation that might be discerned in this inquiry. In view of the fact that previously cited research chronicles links between DRD4 and 5HTTLPR and negative emotionality, (Auerbach et al., 1999; Holmboe et al., 2011) there are certainly grounds for predicting that a G×E effect involving one or both of these polymorphisms could account for a parallel moderational effect involving difficult temperament.

Finally, given the availability of repeated measurements of our dependent variables from just before or during the kindergarten year through sixth grade, we seek to determine whether G×E effects prove to be time limited or endure. Typically, as the time between the measurement of a developmental experience and the developmental outcome it is thought to influence increases, the strength of the association between the two decreases. We can see no reason why this would not apply to a genetically moderated effect of early experience, in this case child-care experience. Thus, we not only seek to determine whether DRD4 moderates the effect of child care around the transition to school, but whether it continues to do so through the elementary school years. We do this by determining whether the child-care-related G×E interactions under investigation predict the slope—or change—in the outcomes studied from initial time of measurement just before school entry (behavior problems) or in kindergarten (social skills) to about 12 years of age, and if so, we then evaluate child-care-related G×E effects on social functioning at subsequent measurement occasions. With the exception of two very recent studies (Petersen et al., 2012; Sulik et al., 2011), virtually all G×E work to date has focused exclusively on outcomes measured at one point in time. Thus, it is impossible to determine whether detected G×E effects dissipate, strengthen, or remain unchanged over time. Unless the possibility that G×E effects change over time is evaluated, the enduring nature of any discerned G×E effect cannot be known.

Method

Participants

Families were recruited through hospital visits to mothers shortly after the birth of a child in 1991 in 10 locations in the United States. During the selected 24-hr intervals, all women giving birth (n = 8,986) were screened for eligibility. From that group, 1,364 families completed a home interview when the infant was 1 month old and became study participants. DNA (buccal cell swabs) was collected when children were 15 years old. Details of the sampling plan can be found in NICHD ECCRN (2005a).

Only children of Caucasian ethnicity were included in this study to avoid confounding effects of ethnic differences in gene frequency. Genetic data were available for 516 of the 1,097 Caucasian children. Of the 516 children with genetic data, 8 were excluded from the analysis sample due to having no outcome data, resulting in a final sample of 508 children. Children excluded, relative to those included, came from households with lower income-to-needs ratios (2.93 vs. 3.89) and with less educated (13.96 vs. 14.70 years) and more depressed (10.27 vs. 9.23) mothers, who were more likely to be single parents (20.80% vs. 9.35%) and who provided lower quality parenting (−.16 vs. .18). Excluded children were rated as having a more difficult temperament in infancy (3.21 vs. 3.13), experienced a lower proportion of center child care prior to starting school (.19% vs. .23%), had more externalizing behavior problems at every assessment from kindergarten (50.30 vs. 49.05) through sixth grade (51.05 vs. 49.32), and had less social skills at every assessment from kindergarten (102.31 vs. 104.81) through sixth grade (101.23 vs. 104.50). Characteristics of the final sample and means of all variables are displayed in Table 1.

Measures

With the exception of DRD4, 5HTTLPR, and difficult temperament, most measures described are very similar to those used in Belsky et al.'s (2007) first evaluation of longer term child-care effects in the NICHD study. Additional information about

Table 1
Sample Characteristics

Variables	*N*	*M* (*SD*)
Maternal education at 1 month [years]	508	14.70 (2.39)
Income-to-needs ratio (across 54 months)	507	3.89 (2.70)
Maternal depression (across 54 months)	508	9.23 (6.18)
Parenting quality (across 54 months)	505	0.18 (0.58)
Partner presence (across 54 months)	508	90.65 (23.16)
Child temperament at 6 months	499	3.13 (0.40)
Child-care quality (across 54 months)	464	2.82 (0.24)
Child-care hours (across 54 months)	502	25.89 (16.07)
Center child care (across 54 months)	495	0.23 (0.28)
Child gender		
Boy	251	49.4%
Girls	257	50.6%
Child DRD4		
DRD4-7R present	135	26.6%
DRD4-7R absent	373	73.4%
Child 5-HTTLPR		
l/l	143	28.1%
l/s	255	50.2%
s/s	110	21.7%
Child externalizing behavior at 54 months	354	49.75 (9.53)
Child externalizing behavior at KG	465	49.05 (8.53)
Child externalizing behavior at G1	471	49.76 (8.32)
Child externalizing behavior at G2	451	49.50 (8.26)
Child externalizing behavior at G3	471	50.40 (8.37)
Child externalizing behavior at G4	454	49.18 (8.21)
Child externalizing behavior at G5	464	49.78 (8.27)
Child externalizing behavior at G6	439	49.32 (8.91)
Child social skills at KG	458	104.81 (13.61)
Child social skills at G1	467	105.06 (13.06)
Child social skills at G2	449	106.86 (13.45)
Child social skills at G3	467	103.84 (13.41)
Child social skills at G4	453	104.16 (13.16)
Child social skills at G5	464	104.23 (13.73)
Child social skills at G6	430	104.50 (14.08)

procedures and measures are provided in *Manuals of Operation* of the study, located at http://www.nichd.nih.gov/research/supported/seccyd.cfm.

Child-Care Characteristics

Nonfamilial child care was defined as regular care by anyone other than the mother, father, or grandparents—including nannies (whether in home or out of home), family day-care providers, and centers. Three aspects of child-care experiences were measured.

Child-care quantity. Parents reported children's hours of routine nonfamilial care during phone and personal interviews conducted at 3- or 4-month intervals from ages 1–54 months, as well as the type(s) of child care used. The *hours* spent in all settings were summed for each of the 17 intervals or "epochs" and parameterized on an hour per week basis. Individual measures of level and rate of change in quantity of care were computed as the individual intercepts and slopes from an unconditional hierarchical linear models (HLM) analysis of these 17 repeated measures. Following NICHD ECCRN (2003), age was centered at the measurement midpoint, 27 months, so the estimated intercept reflected that child's hours per week at 27 months of age.

Child-care type. For each measurement epoch, each of the child's care arrangements was classified as center, child-care home (any home-based care outside the child's own home except care by grandparents), in-home care (any caregiver in the child's own home except father or grandparent), grandparent care, or father care. The proportion of epochs in which the child received care in a center for at least 10 hr per week was used to represent the type of care.

Child-care quality. Multiple observational assessments using the Observational Record of the Caregiving Environment were conducted in the primary child-care arrangement at ages 6, 15, 24, 36, and 54 months to evaluate how sensitive, responsive, stimulating, positive and nonnegative, and nonintrusive caregiving proved to be; for measurement details, see NICHD ECCRN (2002). As with quantity, individual measures of level and change in quality (i.e., slope) were estimated with an unconditional HLM analysis (NICHD ECCRN, 2003).

Covariates

Early childhood. Following Belsky et al. (2007), measures of maternal, child, and family characteristics during infancy and early childhood were collected and used as controls for possible selection bias: *child gender*, *maternal education* (years of schooling at time of child's birth), the proportion of (five measurement) epochs through 54 months in which the mother reported a *husband* or *partner* was present, family income through 54 months calculated as the mean *income-to-needs ratio*, and the mean of *maternal depressive symptoms* assessed by the Center for Epidemiological Studies Depression Scales reported by the mother at 6, 15, 24, 36, and 54 months. These early childhood covariates were included in the reported HLM analyses as time-invariant controls.

Difficult temperament. Temperament was assessed by maternal report at 6 months using an adapted version of the Infant Temperament Questionnaire (Carey & McDevitt, 1978). Items were designed to capture approach, activity, intensity, mood, and adaptability. An overall summary of "difficultness" was calculated with higher values reflecting higher negative emotionality.

Parenting quality 6–54 months. Parenting quality was assessed by (a) maternal sensitivity in semistructured play and (b) observation at home:

1. *Maternal sensitivity*: Mother–child interactions were videotaped in semistructured 15-min observations at 6, 15, 24, 36, and 54 months, with interactional tasks enabling evaluation of age-appropriate qualities of maternal behavior. Videotapes were coded at a central location by raters blind to other information about the families. Intercoder reliability was determined by assigning two coders to 19%–20% of the tapes randomly drawn at each assessment period. Intercoder reliability was calculated as the intraclass correlation coefficient. Reliability for the composite scores used in the current report exceeded .83 at every age. At 6, 15, and 24 months, composite maternal sensitivity scores were created from the sums of three 4-point ratings (maternal sensitivity to child nondistress, intrusiveness [reversed], and positive regard). At 36 and 54 months, the maternal sensitivity composite was the sum of the three 7-point ratings of supportive presence, hostility (reversed), and respect for autonomy. Cronbach's alphas exceeded .70 at every age.
2. The *Home Observation for Measurement of the Environment* (HOME; Caldwell & Bradley, 1984) was administered during home visits at 6, 15, 36, and 54 months. The Infant or Toddler version of the inventory (IT-HOME), composed of 45 items was used across the first 3 years of life. The Early Childhood version of the inventory (EC-HOME), composed of 55 items, was used at 36 and 54 months. A centrally located system of training was used for data collectors at each age. Cronbach's alphas for the total score at each age exceeded .77.

The HOME total and maternal sensitivity ratings were standardized and averaged at each age and then across the first 54 months to create a composite score measure of early parenting quality.

Primary grades. Measures of family demographic and psychological characteristics also were obtained when children were in kindergarten and in first, third, and fifth grades. Following Belsky et al. (2007), the following were included as time-varying covariates in the HLM analyses: *presence of a husband* or *partner* in the household, *income-to-needs ratio, maternal depressive symptoms,* and *parenting quality.* Composite *parenting quality* scores for the primary grades were created similar to parenting quality for 6–54 months (see above) by averaging standardized ratings of observed maternal sensitivity and standardized ratings of observed home environmental quality (HOME; Caldwell & Bradley, 1984), which were assessed at 54 months and first (only maternal sensitivity), third, and fifth grades.

In addition to family-related covariates, *classroom quality* and *after-school experience* during the primary grades were included as controls. Children's classroom experiences were measured using the *classroom observation system for first grade* (Allhusen et al., 2004), *for third grade* (NICHD ECCRN, 2005b), and *for fifth grade* (NICHD ECCRN, 2004). These observations focused on the classroom as well as the specific study child and his or her classroom experiences. Three 7-point global ratings of the classroom environment were made at the end of two (first grade) or eight (third and fifth grades) 44-min observation cycles: overcontrol by teacher, teacher's emotional detachment, teacher's sensitivity to student needs.

Regarding after-school experience, mothers were interviewed by telephone in the fall and spring of kindergarten and first, third, and fifth grades about the study children's out-of-school care. Following Belsky et al. (2007), hours of nonparental out-of-school care arrangements (here named *after-school hours*) were obtained for each school year from the average across the spring and fall reports of the total hours mothers reported across all nonparental out-of-school care arrangements.

Child Outcomes

Externalizing behavior problems. Teachers reported on children's *externalizing behavior problems* (e.g., "hits others," "disobedient at school," "argues a lot") repeatedly—at 54 months, kindergarten, and annually in first, second, third, fourth, fifth, and sixth grades—using the Child Behavior Checklist Teacher Report Form (Achenbach, 1991). Raw scores were converted into standard T scores, based on normative data for children of the same age.

Social skills. Teachers reported on children's *social competence and social skills* (e.g., "makes friends easily," "controls temper when arguing with other children," "asks permission before using someone else's property") repeatedly using the Social Skills

Questionnaire from the Social Skills Rating System (Gresham & Elliott, 1990)—beginning at kindergarten and annually thereafter in first, second, third, fourth, fifth, and sixth grades. For purposes of this report, raw total scores were converted into standardized scores, based on normative data for children of the same age.

Genetic Analyses

DNA was extracted from buccal cell swabs (Freeman et al., 2003). The majority of samples were genotyped twice for both DRD4 ($n = 438$, 86.2%) and 5-HTTLPR ($n = 465$, 91.5%) to evaluate the reliability of genotyping. If there was a discrepancy between the two assessments, genotyping was repeated until the same result was found twice. We defaulted to the original genotype, however, if a sample could not be genotyped a second time or if we were unable to identify a single genotype for a given sample.

DRD4 was identified using a modified assay based on methods developed by Sander et al. (1997) and modified by Anchordoquy, McGeary, Liu, Krauter, and Smolen (2003): 1× Taq Gold Buffer, 2.25 mM final concentration of $MgCl_2$, 10% DMSO, 0.2 mM dNTPs, 0.1 mM deazo GTP, 0.75 μM primers, 40 ng of DNA, and 1 U of Taq Gold (Applied Biosystems, Foster City, CA) in a volume of 12 μl. The primer sequences were: 5′–6-FAM-GCGAC TACGTGGTCTACTCG-3′ and reverse, 5′-AGGACC CTCATGGCCTTG-3′. The amplification procedure was as described by Anchordoquy et al. (2003). One microliter was removed and placed in a 96-well plate and 10 μl of formamide containing LIZ-500 standard (Applied Biosystems). The plate was run using a Fragment Analysis protocol in the 3730XL DNA Analyzer (Applied Biosystems). Fragments were analyzed using Genemapper software (Applied Biosystems) with PCR products of (in bp): 379, 427, 475 (43), 523, 571, 619 (73), 667, 715, 763, and 811. DRD4 was coded as individuals carrying one or more 7-repeat allele versus all others. Agreement between first and second genotyping was 86.5% ($\kappa = .63$, $p < .001$). For the 13.5% where the two rounds of genotyping proved discrepant, a third round was conducted that determined what genotype these discrepant cases would be assigned. The DRD4 7-repeat allele was present in 26.6% of the final sample.

5-HTTLPR was identified using a modified assay based on the method of Lesch et al. (1996) and Anchordoquy et al. (2003): 1× Taq Gold Buffer, 1.8 mM final concentration of MgCl2, 10% DMSO, 0.2 mM dNTPs, 0.1 mM deazo GTP, 0.6 μM primers, 40 ng of DNA, and 1 U of Taq Gold (Applied Biosystems) in a volume of 15 μl. The primer sequences were: forward, 5′-VIC-GGCGT TGCCGCTCTGAATGC-3′ and reverse, 5′-GAGGG ACTGAGCTGGACAACCAC-3′. The same amplification protocol was used as for DRD4 (see above). Fragments were analyzed using Genemapper software (Applied Biosystems) with PCR products of 484 or 528 bp. Agreement between first and second genotyping was 83.7% ($\kappa = .74$, $p < .001$). For the 16.3% where the two rounds of genotyping proved discrepant, a third round was conducted that determined what genotype these discrepant cases would be assigned. Genotype distribution in the final sample (l/l: 28.1%; l/s: 50.2%; s/s: 21.7%) conformed to the Hardy–Weinberg Equilibrium ($p = .98$).

Data Analysis Plan

Data analysis focused on testing the moderating effect of DRD4 and 5-HTTLPR on the long-term associations between child-care experiences during the first 4.5 years and children's externalizing behavior from that age through spring of sixth grade and children's social skills from kindergarten through spring of sixth grade. An analytic strategy similar to that used by Belsky et al. (2007) and Pluess and Belsky (2010) was implemented. HLM (Bryk & Raudenbush, 2002; Singer & Willett, 2003) were fitted to estimate individual and group linear and quadratic growth curves. The models included family and child-care or school experiences measured during both the preschool (i.e., 0–54 months) and primary-school years (i.e., 54 months–sixth grade) as covariates, as well as child DRD4, 5-HTTLPR, and the two-way interactions involving each polymorphism and each of the three child-care variables (i.e., quality, quantity or hours, and center-care experience). Separate models were run to test each possible two-way interaction term between each gene and each child-care variable on both outcomes of interest, yielding a total of 12 separate models. Individual intercepts and linear slopes with respect to age were estimated as correlated random effects for each child. In the interest of space, reporting of results focuses exclusively upon the interactions between DRD4 and early child-care experiences, that is, those effects that extend previous findings (Belsky et al., 2007; Pluess & Belsky, 2010) are of primary interest here and proved statistically significant.

Several modeling decisions were made. First, age was centered at 54 months for externalizing problems and at kindergarten for social skills, as these were the initial occasions at which the problem-behavior and social-skills measures that were repeatedly administered were obtained. Perhaps noteworthy, then, is the fact that these first measurements of the two constructs were provided by different respondents—child-care caregivers at 54 months in the case of behavior problems and kindergarten teachers in the case of social skills. The initial interaction effect coefficients to be presented indicate whether the G×E interactions of interest were related to (a) behavior problems at the age of 54 months or social skills at kindergarten and (b) change—or slope—of externalizing behavior from 54 months to sixth grade or of social skills from kindergarten to sixth grade. If any interaction term predicted the slope, the initial analysis was rerun repeatedly after changing the intercept age so that the changing nature over time of the interaction between the polymorphism and child-care factor in question could be illuminated.

Whenever GXE interactions predicting an intercept proved significant, we conducted a "regions-of-significance" analysis, following Kochanska, Kim, Barry, and Philibert (2011), to evaluate whether the significant interaction proved more consistent with a diathesis-stress or differential-susceptibility model of environmental action. The region of significance defines the specific values of a child-care feature (i.e., quality, quantity, type) at which the slope between a particular polymorphism and externalizing problems moves from significance to nonsignificance and/or vice versa (Aiken & West, 1991; Hayes & Matthes, 2009; Preacher, Curran, & Bauer, 2006). If and when evidence emerged from the regions-of-significance analysis that the data conformed to a differential-susceptibility model of environmental action, we double checked this conclusion by implementing a series of additional and more demanding analytic steps recently outlined by Roisman et al. (2012) for evaluating the nature of the interaction detected.

Missing data occurred in the included sample (n = 508) due to attrition and failure to complete all assessments, as follows with respect to the primary child-care predictors and child-development outcomes: quality of care (8.7%), hours of care (1.2%), center-care experience (2.6%), externalizing behavior (7.3%–30.3%, representing lowest and highest at any one time across repeated measurement occasions), and social skills (8.1%–15.4%). Missing data also occurred in primary-grade covariates, as follows: parenting quality (.6%–5.9%), maternal depression (1.0%–6.9%), income (3.0%–8.7%), partner presence (0.2%–5.5%), classroom quality (3.7%–20.9%), and after-school activities (1.6%–5.3%). Missing data of the included sample were imputed with multiple imputation (Schafer, 1997) using all available data (N = 1,364). Test statistics and regression coefficients were averaged across five imputed data sets. When analyses were run with only cases with complete data, results did not differ from those derived from the imputed data sets. For the simple slopes and regions-of-significance follow-up analyses, intercepts for externalizing behavior and social skills were estimated by means of HLM analyses. These intercepts were then averaged across the five imputed data sets. The level of significance for all analyses was set at $\alpha = .05$.

Results

Four sets of results are presented, the first preliminary, provides simple descriptive statistics on the measurements included in this report, while illuminating the most important bivariate relations among variables; the second primary, addressing the interactions of interest between DRD4 or 5-HTTLPR and child-care experiences in the prediction of children's externalizing behavior and social skills over time; the third, follow-up analyses illuminating the form of significant interactions; and a fourth set evaluating whether the detected G×E interactions remained significant when a previously chronicled interaction involving child care and difficult temperament was taken into account (Belsky & Pluess, 2012; Pluess & Belsky, 2009, 2010).

Preliminary Analysis: Descriptive Statistics and Unadjusted Associations

Table 1 presents the means and standard deviations of the variables of primary interest. With respect to the intercorrelation of these measurements, perhaps what is most important to note is that DRD4 and 5-HTTLPR were not significantly associated with child-care predictors or the child-development outcome variables except as follows: DRD4 with sixth-grade externalizing behavior ($r = -.10$, $p < .05$) and 5-HTTLPR with fifth-grade externalizing behavior ($r = -.09$, $p < .05$) and social skills ($r = .10$, $p < .05$). Given the lack of significant associations between genes and outcomes at all other assessment points, the data appear to fulfill the criteria of independence of the moderator variable (i.e., DRD4, 5-HTTLPR) from the environment (i.e. child care) and outcome (i.e., externalizing problems, social skills), thereby generally meeting criteria for testing differential susceptibility (Belsky, Bakermans-Kranenburg, et al., 2007). Externalizing behavior and social skills were significantly and negatively associated at all assessments (range $r = -.13$ to $-.60$, $p < .01$).

Primary Analyses

For the primary analyses predicting, first, externalizing behavior and, second, social skills, we ran HLM across the multiple assessment points. Variables included child DRD4 and 5-HTTLPR, child-care quality intercept (estimated quality of care at 27 months, reflecting the midpoint between 6 and 54 months), the hours per week intercept (estimated from HLM analyses in which the intercept was set at 27 months), proportion of 3- to 4-month epochs in center-based child care for at least 10 hr per week, six preschool time-invariant covariates (child gender, infant temperament at 6 months, maternal education at month 1, mean income-to-needs ratio between 1 and 54 months, mean parenting quality between 6 and 54 months, mean maternal depressive symptoms between 6 and 54 months), and five concurrent time-varying covariates from 54 months through sixth grade (income-to-needs ratio, parenting quality, maternal depression, observed school classroom quality, and hours per week of after-school care (set to 0 for 54 months). The model was run once including only the covariates, child-care variables, and the two polymorphisms, all as main effects, and then again with two-way interactions involving DRD4 and 5-HTTLPR with each of the three child-care variables (i.e., six separate models for each outcome). Because 5-HTTLPR did not significantly predict the intercept or slope of either externalizing behavior or social skills, whether as a main effect or in interaction with any of the child-care variables, reporting below pertaining to genetic findings is restricted to those involving DRD4; Tables 2 and 3 present results for externalizing problems and social skills, respectively.

Externalizing Behavior

With regard to main effects, findings displayed in Table 2 indicate that externalizing problems at 54 months were significantly greater when parenting quality across the early years was lower, when children lived in a single-parent home (i.e., less partner presence), when children were exposed to higher proportions of center care, and when they spent more hours in child care irrespective of type of care. The amount of time spent in child care also significantly predicted the decline in externalizing behavior across the primary grade years. None of the time-variant primary-grade covariates significantly predicted the intercept or slope of externalizing behavior problems except after-school experience, which predicted both the intercept ($B = -0.06$, 95% CI [-0.12, -0.003], $p = .04$) and the slope ($B = 0.03$, 95% CI [0.01, 0.05], $p < .01$). Notably, neither DRD4 nor child-care quality significantly predicted problem behavior (intercept or slope) as main effects.

DRD4 and child-care quality did interact, however, to significantly predict externalizing behavior centered at 54 months ($B = -6.23$, 95% CI [-12.26, -0.21], $p = .04$) and to marginally predict the slope of externalizing behavior across the eight assessment points ($B = 0.96$, 95% CI [-0.11, 2.03], $p = .08$). The latter result was followed up below to determine whether the G×E in question remained significant at times of measurement after 54 months of age. The slope of externalizing problems was also significantly predicted by interactions involving DRD4 and child-care hours ($B = -0.02$, 95% CI [-0.04, -0.003], $p = .02$) and child-care type ($B = -0.97$, 95% CI [-1.88, -0.06], $p = .04$), but because follow-up analyses never revealed significant relations between these interactions and externalizing intercepts at any of the ages of measurement, in the interest of space, details of such analyses are not reported.

Social Skills

Results pertaining to main effects indicated that social skills at kindergarten were significantly greater when parenting quality across the early years was higher and when children did not live in a single-parent home (i.e., more partner presence). None of the time-variant primary-grades' covariates significantly predicted the intercept or slope of social skills except after-school experience, which predicted the slope ($B = -0.04$, 95% CI [-0.08, -0.005], $p = .02$). Notably, neither DRD4 nor any of the child-care variables predicted the social skills intercept at kindergarten as main effects.

DRD4 and child-care quality did interact, however, to predict social skills centered at kindergarten ($B = 10.67$, 95% CI [1.23, 20.11], $p = .03$)

Table 2
Summary of Hierarchical Linear Model Predicting Behavior Problems (N = *508)*

Predictor variables	Behavior problems intercept centered at 54 months *B*	Behavior problems slope *B*
Step 1		
Maternal education at 1 month	0.06	−0.04
Income-to-needs ratio (across 54 months)	0.18	−0.05
Maternal depression (across 54 months)	−0.03	< −0.01
Parenting quality (across 54 months)	−2.63**	0.03
Partner presence (across 54 months)	−0.05**	< 0.01
Child temperament at 6 months	−0.17	−0.18
Child gender (1 = *male*; 2 = *female*)	−0.02	0.02
Child-care quality (across 54 months)	0.02	0.37
Child-care hours (across 54 months)	0.08**	−0.01**
Center child care (across 54 months)	4.00**	−0.28
Child DRD4 (0 = *7R absent*, 1 = *7R present*)	0.59	−0.21
Child 5-HTTLPR (0 = l/l; 1 = l/s; 2 = s/s)	−0.15	−0.18
Step 2		
DRD4 × Child-Care Quality	−6.23*	0.96†
DRD4 × Child-Care Hours	0.07	−0.02*
DRD4 × Center Child Care	2.66	−0.97*

Note. The model included the following time-variant covariates from 54 months through sixth grade that are not displayed in the table: income-to-needs ratio, maternal depression, parenting quality, partner presence, classroom quality, after-school activities. The displayed coefficients of the variables at Step 1 represent the values before inclusion of interaction terms at Step 2.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

Table 3
Summary of Hierarchical Linear Model Predicting Social Skills (N = *508)*

Predictor variables	Social skills intercept centered at kindergarten *B*	Social skills slope *B*
Step 1		
Maternal education at 1 month	0.49†	−0.04
Income-to-needs ratio (across 54 months)	−0.15	0.08
Maternal depression (across 54 months)	0.03	< 0.01
Parenting quality (across 54 months)	3.60*	0.03
Partner presence (across 54 months)	0.06*	< 0.01
Child temperament at 6 months	0.81	0.15
Child gender (1 = *male*; 2 = *female*)	1.34	−0.40†
Child-care quality (across 54 months)	1.85	−0.12
Child-care hours (across 54 months)	−0.04	0.01
Center child care (across 54 months)	−3.25	0.73
Child DRD4 (0 = *7R absent*; 1 = *7R present*)	−0.46	0.18
Child 5-HTTLPR (0 = l/l; 1 = l/s; 2 = s/s)	−0.02	0.13
Step 2		
DRD4 × Child-Care Cuality	10.67*	−2.48*
DRD4 × Child-Care Hours	−0.01	< 0.02
DRD4 × Center Child Care	5.45	−0.82

Note. The model included the following time-variant covariates from kindergarten through sixth grade that are not displayed in the table: income-to-needs ratio, maternal depression, parenting quality, partner presence, classroom quality, after-school activities. The displayed coefficients of the variables at Step 1 represent the values before inclusion of interaction terms at Step 2.
$^{\dagger}p < .10$. $^{*}p < .05$.

and the slope of social skills across the eight assessment points ($B = -2.48$, 95% CI [−4.52, −0.44], $p = .02$). The latter result was followed up below to determine whether the G×E in question remained significant at times of measurement after kindergarten.

Secondary and Tertiary Analyses: Form and Timing of G×E Interaction

Two sets of follow-up analyses were carried out to illuminate the significant DRD4 × Child-Care Quality interactions, one pertaining to the form of the interaction and the other to its potentially changing nature over time.

Interaction Form

To interpret the significant DRD4 × Child-Care Quality interaction in predicting the externalizing-behavior intercept at 54 months and the social-skills intercept at kindergarten, we plotted regression slopes of child-care quality on predicted externalizing problems at 54 months and social skills at kindergarten, separately for children with and without the DRD4 7-repeat allele. Figure 1 indicates that the relation between child-care quality and externalizing problems was negative and significant in the case of children carrying the DRD4 7-repeat allele ($\beta = -.27$, $p < .01$), but not in the case of children without the DRD4 7-repeat ($\beta = -.07$, $p = .21$). Figure 2 indicates that the relation between child-care quality and social skills was positive and significant in the case of children carrying the DRD4 7-repeat allele ($\beta = .35$, $p < .01$), but not in the case of children without the DRD4 7-repeat ($\beta = .10$, $p > .05$).

Visual inspection of Figures 1 and 2 reveals a crossover interaction consistent with differential-susceptibility in that children carrying the DRD4 7-repeat allele had the most externalizing problems when exposed to low-quality child care early in life yet the least such problems when quality was high, as well as the most social skills when quality was high and the least when quality was low. Analysis of regions of significance of the data on which Figure 1 is based, using a tool provided by Preacher et al. (2006), yielded only a lower bound of significance within the observed range of child-care quality, however; that is, the slope between DRD4 and externalizing problems proved significant when child-care quality was lower than 2.45 (i.e., shaded areas in Figure 1), representing 6.5% of the sample, with no significant differences emerging above this value. Analysis of the data on which Figure 2 is based, on the other hand, yielded both a lower and a higher bound of significance within the observed range of child-care quality; more specifically, the slope between DRD4 and social skills proved significant when child-care quality was lower than 2.53, representing 9.3% of the sample, and higher than 3.24, representing 2.2% of the sample (i.e., shaded areas in Figure 2). Consequently, the significant interaction between DRD4 and child-care quality in the prediction of externalizing problems at 54 months

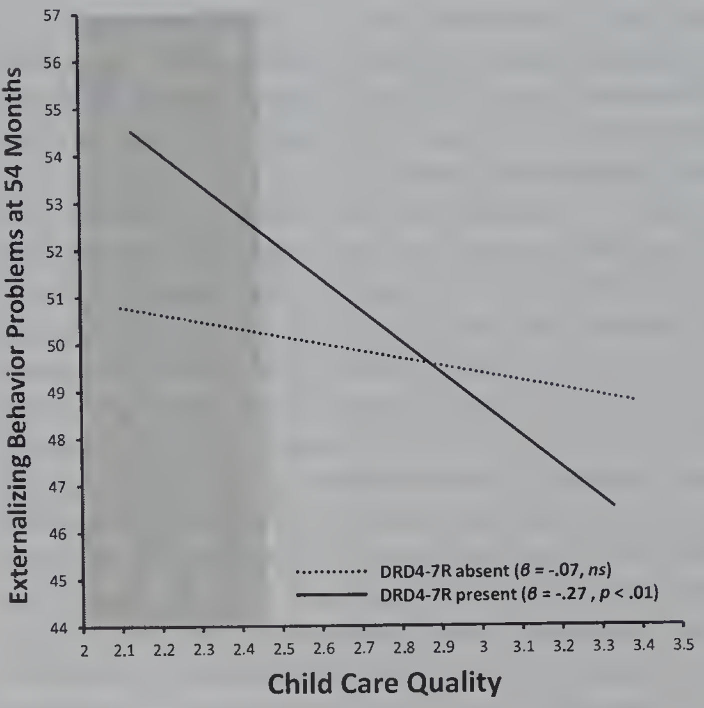


Figure 1. DRD4 × Child-Care Quality interaction predicting teacher-reported externalizing problems at 54 months. The shaded area represents the region of significance.

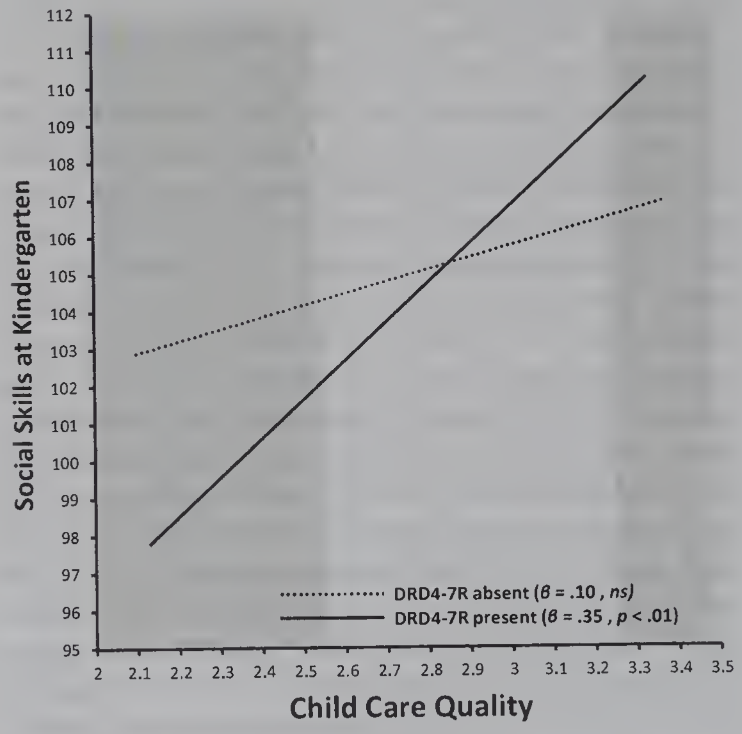


Figure 2. DRD4 × Child-Care Quality interaction predicting teacher-reported social skills at kindergarten. The shaded area represents the region of significance.

was more consistent with a diathesis-stress model of environmental action, whereas the significant interaction between DRD4 and social skills at kindergarten was more consistent with a differential-susceptibility model.

Before confidently embracing a differential-susceptibility conclusion with respect to social skills at kindergarten, we proceeded to implement a series of additional analyses that Roisman et al. (2012) proposed for evaluating differential-susceptibility (using an Internet tool: http://www.yourpersonality.net/interaction/ros2.pl). According to these additional analyses, the crossover point of the simple slopes on child-care quality was at 2.85, well within their criterion range of ± 2 *SD* (*SD* = .23) from the mean (*M* = 2.82). The proportion of the interaction (PoI) provides a way to express the proportion of the total interaction that is represented below and above the crossover point, with PoIs close to .50 regarded by Roisman et al. as strong evidence of differential susceptibility. PoI in the ± 2 *SD* range with respect to social skills at kindergarten was .57 below and .43 above the crossover point, clearly close to the .50 criteria. The proportion affected (PA) index represents the proportion of cases that fall above the crossover point with larger percentages suggesting stronger evidence for differential-susceptibility relative to diathesis stress. PA with respect to social skills was .49. Hence, these additional analyses suggest that the interaction between child-care quality and DRD4 in the prediction of social skills does reflect a differential-susceptibility rather than a diathesis-stress pattern of environmental action, consistent with the visual interpretation of the regions-of-significance analysis.

G×E interaction over time. To illuminate the nature of the significant interaction between DRD4 and child-care quality in predicting the slope of behavior problems and of social skills, we ran, respectively, six or seven additional HLM models, each centered at a different assessment point (externalizing: kindergarten and first through sixth grade; social skills: first through sixth grade). After being significant at 54 months (see above), the interaction between DRD4 and child-care quality in the prediction of behavior problems proved marginally significant at kindergarten (*B* = –5.35, 95% CI [–10.83, 0.13], *p* = .06) and nonsignificant thereafter (*B*s = –4.00 to 0.68, *p* > .10 for all). Similarly, after being significant at kindergarten (see above), the interaction between DRD4 and child-care quality in the prediction of social skills proved marginally significant at first grade (*B* = 7.45, 95% CI [–0.56, 15.46], *p* = .07) and nonsignificant thereafter (*B*s = 4.89 to – 4.90, *p* > .10 for all). In both cases, then, the data suggest that the genetic moderation of effects of early child-care quality on children's social-emotional adjustment dissipated over time.

Including Child-Care Quality × Difficult Temperament Interaction

Recall that in earlier analyses of NICHD study data we discerned significant interactions between infant temperament and child-care quality in predicting behavior problems and social skills (Belsky & Pluess, 2012; Pluess & Belsky, 2009, 2010). Given the aforementioned evidence indicating that DRD4 is related to infant negative emotionality (Holmboe et al., 2011; Ivorra et al., 2010), we sought to determine whether the DRD4 × Child-Care Quality interactions reported herein were a function of the previously detected Temperament × Child-Care Quality interaction effects. Thus, we reran the two HLM with the significant interaction term between DRD4 and child-care quality predicting (a) externalizing behavior problems at 54 months and (b) social skills at kindergarten with an additional two-way interaction term involving difficult infant temperament and child-care quality added to the model (along with the main effect of temperament). Consistent with earlier findings (Belsky & Pluess, 2012; Pluess & Belsky, 2009, 2010), the interaction between temperament and child-care quality predicting the 54-month problem-behavior intercept proved significant (*B* = –10.68, 95% CI [–17.79, –3.57], *p* < .01), while the G×E interaction between DRD4 and child-care quality remained marginally significant (*B* = –5.50, 95% CI [–11.48, 0.47], *p* = .07). Similarly, the interaction between temperament and child-care quality predicting the kindergarten social-skills intercept proved marginally significant (*B* = 10.82, 95% CI [–1.04, 22.69], *p* = .07), whereas the G×E interaction between DRD4 and child-care quality remained significant (*B* = 9.96, 95% CI [0.51, 19.41], *p* = .04). These findings suggest that that the moderating effects of infant temperament and DRD4 vis-à-vis child-care quality and its effects on problem behavior and social skills are largely, even if not entirely, independent.

Discussion

The research reported herein had multiple goals, each of which is considered in turn before turning to study limitations.

G×E Interaction and Its Form

In this first study investigating whether effects of day care on children's social adjustment across the childhood years—reflected in their externalizing problems and social skills—might be genetically moderated, evidence emerged that this was indeed the case, but only with regard to DRD4, not 5-HTTLPR and quality of care. Recall that DRD4 and 5-HTTLPR were selected for inclusion in this research because it is these two polymorphisms for which the most G×E evidence has emerged seemingly consistent with differential susceptibility. It should be appreciated, however, that this could itself be an artifact of these two polymorphisms being among the most studied from a G×E perspective. It was somewhat surprising that no significant G×E interactions emerged in the case of 5-HTTLPR given the results of a recent meta-analysis evaluating—and documenting—such effects in the case of children under 18 years of age; indeed, it provided evidence of G×E consistent with differential susceptibility (van IJzendoorn et al., in press). It did not include, however, any research evaluating child-care effects.

With regard to the G×E effects detected, these involved DRD4 and quality of care predicting externalizing problems at 54 months and kindergarten and social skills at kindergarten and first grade. One of the things that makes such findings noteworthy is that different respondents provided evaluations of child behavior across the years of measurement covered in this inquiry: caregivers at 54 months, kindergarten teachers during the 1st year of school, first-grade teachers during the 2nd year, and so on through sixth grade.

The graphical plotting of the detected G×E interactions highlighting a crossover-interaction pattern appeared more consistent with differential-susceptibility rather than diathesis-stress thinking. The regions-of-significance analysis indicated otherwise, however, in the case of externalizing problems. Recall that it was only at low levels of quality of care that differences between children with and without the 7-repeat allele manifested themselves, with those carrying the allele appearing more adversely affected by low-quality care (i.e., manifesting more behavior problems), at least during the first 2 years that each outcome was measured, than those not carrying this allele. The fact that the regions-of-significance analysis also revealed that children carrying the 7-repeat allele who were exposed to higher quality care benefited more than other children from such supportive environments when the outcome to be explained was social skills proved consistent with the differential-susceptibility hypothesis. That is, there was evidence of both "for-better-*and*-for-worse" effects of child-care quality on children carrying the 7-repeat allele.

Of note, however, is that the percentage of children exposed to quality of care at a level that yielded a reliable developmental benefit—as revealed by the regions-of-significance analysis (see Figure 2)—was relatively small (2.2%), clearly raising questions about just how strong was the "for better" side of the differential-susceptibility equation in this instance. Nevertheless, when the crossover interaction indicative of differential-susceptibility was evaluated further by applying Roisman et al.'s (2012) evidentiary standards, the data continued to highlight the differential-susceptibility nature of the interaction under consideration.

Beyond Main Effects

On the basis of the findings just summarized, it would seem that prior analyses by the NICHD ECCRN of child-care effects and, in particular, quality-of-care effects on social adjustment from 54 months through sixth grade failed to detect any such effects because the original investigators only focused on main effects or effects moderated by child gender and family risk conditions (Belsky et al., 2007; NICHD ECCRN, 2003). Apparently—given the nonexperimental nature of the NICHD Study—low-quality child care does matter with respect to problem behavior and social skills around the time of the transition to school, but only for some children, not others, with the same perhaps being true of high-quality care in the case of social skills. If such findings can be replicated, they would call into question the widespread presumption that poor quality of care compromises the development of most children, at least with respect to socioemotional functioning, and that good quality care yields measurable benefits in this realm of functioning for most children. But even if this proves to be the case in future work, it would not provide grounds for reducing investment in quality of child care. Not only does the NICHD study and many others consistently find quality of care to (modestly) predict cognitive and language development (NICHD ECCRN, 2006), as well as academic achievement (Belsky et al., 2007; Vandell et al., 2010), but humanitarian considerations alone dictate the provision of good quality care conditions for all children.

The Moderating Effect of Difficult Temperament

Because previous analyses of NICHD study data indicated that quality effects on problem behavior or social skills were moderated by difficult infant temperament, with infants scoring high in negativity proving more susceptible—for-better-*and*-for-worse—to the apparent influence of child-care quality than other infants, the issue arose as to whether the G×E interactions discerned in this inquiry involving DRD4 could underlie—and account for—the parallel interactions involving temperament reported by Pluess and Belsky (2009, 2010). As it turned out, this did not prove to be the case. Recall that in the final set of analyses presented, both interactions involving quality of care proved significant or marginally so. This means that (a) infants with difficult temperaments or (b) children carrying the 7-repeat DRD4 allele are disproportionately, even exclusively, susceptible to effects of low-quality care on externalizing problem behavior and social skills around the transition to school.

To our knowledge, this is the first demonstration in a single inquiry of independent interactions involving temperament and genetics with the same (or even a different) environmental factor. There would seem to be reason to question, then, any conceptual privilege granted genes over temperament (or physiology; Boyce & Ellis, 2005) when it comes to considering organismic factors that might moderate environmental effects on human development, either in diathesis-stress or differential-susceptibility terms. It should be acknowledged, nevertheless, that results pertaining to the issue of overlap between interactions between child-care quality and temperament and DRD4 on which this conclusion is based might have been different had other measures of temperament, perhaps ones based on observations rather than maternal reports, been subject to analysis. Only future research will be in position to address this possibility.

Developmental Analysis

Besides focusing on the genetic moderation of child-care effects and distinguishing differential-susceptibility from diathesis-stress models of environmental action, one of the strengths of the current inquiry was that it was developmental in character. Because externalizing problems and social skills were repeatedly measured—by independent evaluators (i.e., different teachers each academic year)—we were in position not just to determine whether child-care effects were moderated by the child's genetic makeup at some point in time, but whether the G×E effect initially detected in problem behavior at 54 months and in social skills at kindergarten endured, dissipated, or even strengthened over time. This we could do by predicting in the initial model the slope of externalizing problem over time, that is, change in externalizing problems, and then shifting the intercept predicted in subsequent rounds of analysis from 54 months to kindergarten to first grade all the way through sixth grade.

As it turned out, the G×E interactions proved significant at the initial occasion at which the outcome was measured, but not thereafter. This seems likely to have been the result of experiences—in the family, at school, and in after-school programs perhaps—that intervened between when child-care experiences and later assessments of child functioning were obtained. Clearly, development is not over by the time children start school.

This observation highlights the need to avoid the implicit presumption that a G×E interaction (of any kind) detected at one point in time, as is the typical practice in most G×E inquiry, necessarily endures over time. At the same time, investigators should be alert to the possibility, not discerned in this inquiry, that a failure to detect such an interaction at one point in time does not automatically imply that it could not emerge subsequently. Only by considering functioning at multiple points in time and examining G×E effects on slope or change over time in a developmental outcome of interest will it be possible to address these matters.

Study Limitations

Perhaps the two biggest limitations of the current inquiry are that it was nonexperimental in nature and included only 508 of some 1,364 children initially enrolled in the NICHD study. The observational character of this investigation clearly limits any and all causal inferences that can be drawn from the findings presented. Nevertheless, language pertaining to child-care "effects" was employed throughout this report for heuristic purposes. A further, yet minor, limitation is that the study did not genotype SNP rs25531 and differentiate between L_A and L_G alleles, and thus we were not positioned to treat the L_G alleles as short alleles (Hu et al., 2006). One can also wonder whether, had the genotyping proven more reliable, more evidence of G×E interactions might have emerged, especially after the initial time of outcome measurements, or whether they might have proven more consistent with differential susceptibility rather than diathesis stress.

With regard to the latter, it also seems possible that greater inclusion of exceptionally high-quality child-care experiences might have resulted in findings more consistent with differential susceptibility.

The limited number of research participants included in this research was mostly due to the fact that not all families agreed to contribute DNA on their child and that 2 of the 10 participating research sites could not secure ethics' approval to request DNA from study participants' families. There was also a need, based on racial differences in gene frequencies, to keep the sample racially homogenous. Due to the consequential substantial reduction in sample size, it is difficult to be sure that the results presented herein would generalize to the original sample, to say nothing of a nationally—or internationally—representative one. Especially important in this regard is that the NICHD study children not included in this inquiry differed from those who were included in terms of their child-care experience and their social functioning.

In some sense, then, the results of this study should be regarded more as "proof of concept" rather than a basis for definitive conclusions. Results suggest that at least some child-care effects may vary across children as a result of their genetic makeup. They further provide evidence that G×E effects detected at one point in time may or may not endure over time. They additionally indicate that moderational effects involving temperament or any other child characteristic should not be presumed to reflect some underlying genetic factor, at least not without evidence that the moderating effect of a temperamental factor is accounted for by the moderating effect of a genetic one.

References

Achenbach, T. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Allhusen, V., Belsky, J., Booth-LaForce, C. L., Bradley, R., Brownwell, C. A., Burchinal, M., et al. (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology, 40*, 651–664.

Anchordoquy, H. C., McGeary, C., Liu, L., Krauter, K. S., & Smolen, A. (2003). Genotyping of three candidate genes after whole-genome preamplification of DNA collected from buccal cells. *Behavior Genetics, 33*, 73–78.

Auerbach, J., Geller, V., Lezer, S., Shinwell, E., Belmaker, R. H., Levine, J., et al. (1999). Dopamine D4 receptor (D4DR) and serotonin transporter promoter (5-HTTLPR) polymorphisms in the determination of temperament in 2-month-old infants. *Molecular Psychiatry, 4*, 369–373.

Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2006). Gene-environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers. *Developmental Psychobiology, 48*, 406–409.

Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Differential susceptibility to rearing environment depending on dopamine-related genes: New evidence and a meta-analysis. *Development and Psychopathology, 23*, 39–52.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Pijlman, F. T., Mesman, J., & Juffer, F. (2008). Experimental evidence for differential susceptibility: Dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Developmental Psychology, 44*, 293–300.

Barry, R. A., Kochanska, G., & Philibert, R. A. (2008). G × E interaction in the organization of attachment: Mothers' responsiveness as a moderator of children's genotypes. *Journal of Child Psychology and Psychiatry, 49*, 1313–1320.

Beaver, K. M., Wright, J. P., DeLisi, M., Walsh, A., Vaughn, M. G., Boisvert, D., et al. (2007). A gene × gene interaction between DRD2 and DRD4 is associated with conduct disorder and antisocial behavior in males. *Behavioral and Brain Functions, 3*, 30.

Belsky, J. (1986). Infant day care: A cause for concern? *Zero to Three, 7*, 1–7.

Belsky, J. (1988). The "effects" of infant day care reconsidered. *Early Childhood Research Quarterly, 3*, 235–272.

Belsky, J. (1990). Parental and nonparental care and children's socioemotional development: A decade in review. *Journal of Marriage and Family, 52*, 885–903.

Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). For better and for worse: Differential Susceptibility to environmental influences. *Current Directions in Psychological Science, 16*, 300–304.

Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry, 14*, 746–754.

Belsky, J., & Pluess, M. (2009). Beyond diathesis-stress: Differential susceptibility to environmental influences. *Psychological Bulletin, 135*, 885–908.

Belsky, J., & Pluess, M. (2012). Differential susceptibility to long-term effects of quality of child care on externalizing behavior in adolescence? *International Journal of Behavioral Development, 36*, 2–10.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T. (2007). Are there long-term effects of early child care? *Child Development, 78*, 681–701.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*, 271–301.

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Caldwell, B., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas Press.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the Infant Temperament Questionnaire. *Pediatrics*, *61*, 735–739.

Caspi, A., Hariri, A. R., Holmes, A., Uher, R., & Moffitt, T. E. (2010). Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American Journal of Psychiatry*, *167*, 509–527.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851–854.

Caspi, A., & Moffitt, T. E. (2006). Gene-environment interactions in psychiatry: Joining forces with neuroscience. *Nature Reviews Neuroscience*, *7*, 583–590.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*, 386–389.

Clarke-Stewart, K. A. (1989). Infant day care. Maligned or malignant? *American Psychologist*, *44*, 266–273.

Crockenberg, S. C. (2003). Rescuing the baby from the bathwater: How gender and temperament (may) influence how child care affects child development. *Child Development*, *74*, 1034–1038.

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049.

Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary–neurodevelopmental theory. *Development and Psychopathology*, *23*, 7–28.

Fox, N., & Fein, G. (Eds.) (1990). *Infant day care: The current debate*. Norwood, NJ: Ablex.

Freeman, B., Smith, N., Curtis, C., Huckett, L., Mill, J., & Craig, I. W. (2003). DNA from buccal swabs recruited by mail: Evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behavior Genetics*, *33*, 67–72.

Gresham, F., & Elliott, S. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavioral Research Methods*, *41*, 924–936.

Holmboe, K., Nemoda, Z., Fearon, R. M., Sasvari-Szekely, M., & Johnson, M. H. (2011). Dopamine D4 receptor and serotonin transporter gene effects on the longitudinal development of infant temperament. *Genes, Brain, and Behavior*, *10*, 513–522.

Howes, C. (1988). Relations between early child care and schooling. *Developmental Psychology*, *24*, 53–57.

Hu, X. Z., Lipsky, R. H., Zhu, G., Akhtar, L. A., Taubman, J., Greenberg, B. D., et al. (2006). Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. *American Journal of Human Genetics*, *78*, 815–826.

Ivorra, J. L., Sanjuan, J., Jover, M., Carot, J. M., Frutos, R., & Molto, M. D. (2010). Gene-environment interaction of child temperament. *Journal of Developmental and Behavioral Pediatrics*, *31*, 545–554.

Karen, R. (1998). *Becoming attached: First relationships and how they shape our capacity to love*. Oxford, England: Oxford University Press.

Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evidence of genetic moderation. *Archives of General Psychiatry*, *68*, 444–454.

Kochanska, G., Kim, S., Barry, R. A., & Philibert, R. A. (2011). Children's genotypes interact with maternal responsive care in predicting children's competence: Diathesis-stress or differential susceptibility? *Development and Psychopathology*, *23*, 605–616.

Langlois, J. H., & Liben, L. S. (2003). Child care research: An editorial perspective. *Child Development*, *74*, 969–1226.

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., et al. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, *274*, 1527–1531.

Maccoby, E. E., & Lewis, C. C. (2003). Less day care or different day care? *Child Development*, *74*, 1069–1075.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*, 133–164.

NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 1/2? *Developmental Psychology*, *39*, 451–469.

NICHD Early Child Care Research Network. (2004). *COS qualitative scales and reliabilities observational ratings of the fifth grade classroom*. (Child Care Data Report 571). Retrieved March 15, 2005, from https://secc.rti.org/project/display.cfm?t=c&i=ccdr571

NICHD Early Child Care Research Network. (2005a). *Child care and child development: Results of the NICHD Study of Early Child Care and Youth Development*. New York: Guilford.

NICHD Early Child Care Research Network. (2005b). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *Elementary School Journal*, *105*, 305–323.

NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist*, *61*, 99–116.

Oades, R. D., Lasky-Su, J., Christiansen, H., Faraone, S. V., Sonuga-Barke, E. J., Banaschewski, T., et al. (2008). The influence of serotonin- and other genes on impulsive behavioral aggression and cognitive impulsivity in children with attention-deficit/hyperactivity disorder (ADHD): Findings from a family-based association test (FBAT) analysis. *Behavioral and Brain Functions*, *4*, 48.

Peisner-Feinberg, E., & Burchinal, M. (1997). Concurrent relations between child care quality and child outcomes: The study of cost, quality and outcomes in child care centers. *Merrill-Palmer Quarterly*, *43*, 451–477.

Petersen, I. T., Bates, J. E., Goodnight, J. A., Dodge, K. A., Lansford, J. E., Pettit, G. S., et al. (2012). Interaction between serotonin transporter polymorphism (5-HTTLPR) and stressful life events in adolescents' trajectories of anxious/depressed symptoms. *Developmental Psychology*, *48*, 1463–1475.

Phillips, D., Fox, N. A., & Gunnar, M. (2011). Same place, different experiences: Bringing individual differences to research in child care. *Child Development Perspectives*, *5*, 44–49.

Phillips, D., McCartney, K., Scarr, S., & Howes, C. (1987). Selective review of infant day care research: A cause for concern. *Zero to Three*, *7*, 18–21.

Pluess, M., & Belsky, J. (2009). Differential susceptibility to rearing experience: The case of childcare. *Journal of Child Psychology and Psychiatry*, *50*, 396–404.

Pluess, M., & Belsky, J. (2010). Differential susceptibility to parenting and quality child care. *Developmental Psychology*, *46*, 379–390.

Pluess, M., Velders, F. P., Belsky, J., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Jaddoe, V. W., et al. (2011). Serotonin transporter polymorphism moderates effects of prenatal maternal anxiety on infant negative emotionality. *Biological Psychiatry*, *69*, 520–525.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regressions, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448.

Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., et al. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *Journal of the American Medical Association*, *301*, 2462–2471.

Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis-stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*, *24*, 389–409.

Rutter, M., Thapar, A., & Pickles, A. (2009). Gene-environment interactions: Biologically valid pathway or artifact? *Archives of General Psychiatry*, *66*, 1287–1289.

Sander, T., Harms, H., Dufeu, P., Kuhn, S., Rommelspacher, H., & Schmidt, L. G. (1997). Dopamine D4 receptor exon III alleles and variation of novelty seeking in alcoholics. *American Journal of Medical Genetics*, *74*, 483–487.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schmidt, L. A., Fox, N. A., Rubin, K. H., Hu, S., & Hamer, D. H. (2002). Molecular genetics of shyness and aggression in preschoolers. *Personality and Individual Differences*, *33*, 227–238.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Sulik, M. J., Eisenberg, N., Lemery-Chalfant, K., Spinrad, T. L., Silva, K. M., Eggum, N. D., et al. (2011). Interactions between serotonin transporter gene haplotypes and quality of mothers' parenting predict the development of children's noncompliance. *Developmental Psychology*, *48*, 740–754.

Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. *Molecular Psychiatry*, *15*, 18–22.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of Early Child Care and Youth Development. *Child Development*, *81*, 737–756.

Vandell, D. L., Henderson, V. K., & Wilson, K. S. (1988). A longitudinal study of children with day-care experiences of varying quality. *Child Development*, *59*, 1286–1292.

van IJzendoorn, M. H., Belsky, J., & Bakermans-Kranenburg, M. J. (in press). Serotonin transporter genotype 5HTTLPR as a marker of differential susceptibility? A meta-analysis of child and adolescent gene-by-environment studies. *Translational Psychiatry*.

Zuckerman, M. (1999). *Vulnerability to psychopathology: A biosocial model*. Washington, DC: American Psychological Association.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1226–1240

Coparenting Among Lesbian, Gay, and Heterosexual Couples: Associations With Adopted Children's Outcomes

Rachel H. Farr
University of Massachusetts Amherst

Charlotte J. Patterson
University of Virginia

Coparenting is associated with child behavior in families with heterosexual parents, but less is known about coparenting among lesbian- and gay-parent families. Associations were studied among self-reported divisions of labor, coparenting observations, and child adjustment (M_{age} = 3 years) among 104 adoptive families headed by lesbian, gay, or heterosexual couples. Lesbian and gay couples reported sharing child care, whereas heterosexual couples reported specialization (i.e., mothers did more child care than fathers). Observations confirmed this pattern—lesbian and gay parents participated more equally than heterosexual parents during family interaction. Lesbian couples showed the most supportive and least undermining behavior, whereas gay couples showed the least supportive behavior, and heterosexual couples the most undermining behavior. Overall, supportive coparenting was associated with better child adjustment.

Family systems theory suggests that children's development cannot be fully understood without considering the network of relationships influencing them (Feinberg, 2003). It is well established that parent–child and marital relationships influence child outcomes, but researchers have more recently begun examining other family relationships (Bornstein & Sawyer, 2006). Coparenting, sometimes called "the family's executive subsystem," refers to the degree of coordination between two adults in their roles as parents (McHale & Irace, 2011), and it has been found to be more strongly tied to child adjustment than other aspects of the couple relationship (Feinberg, 2003). Coparenting has been studied among heterosexual couples and their biological children, but rarely among adoptive families or families with lesbian and gay parents (Patterson & Farr, 2011). As increasing numbers of families in the United States are not composed of a heterosexual mother and father rearing biological children (e.g., Patterson & Hastings, 2007), it is critical to understand child development in the context of other family systems.

Coparenting includes the ways in which couples carry out the tasks of parenting (e.g., divisions of child-care labor) and discrepancies in parental investment, as well as both overt and covert parental behaviors that could be supportive or undermining (McHale, Kuersten-Hogan, & Lauretti, 2001). Regardless of sexual orientation, two adults who parent together must manage the many tasks associated with rearing children. Thus, coparenting involves coordination between the two adults to complete such tasks. Associations between coparenting and child development are well documented in heterosexual-parent families with biological children (McHale & Irace, 2011), particularly young children (e.g., Schoppe, Mangelsdorf, & Frosch, 2001; Stroud, Durbin, Wilson, & Mendelsohn, 2011). Variations in coparenting contribute uniquely to children's development, beyond the influence of marital quality and individual parent–child relationships (Teubert & Pinquart, 2010). As most research on coparenting has focused on heterosexual couples and their children, and generally on the biological children of these couples (McHale & Irace, 2011), more research is needed on coparenting and its associations with child outcomes in diverse families. In particular, the extent to which coparenting dynamics is influenced by parent gender, sexual orientation, and biological parent–child relationships

---

This research was supported in part by grants from the Williams Institute at UCLA and from the Lesbian Health Fund to Charlotte J. Patterson. We also wish to thank: Camila Escobar, Candace Garramone, Stacy Kruczkowski, Johannah Merrill, Alexandria Moore, and Kathryn Trizna for their contributions to this research.

Correspondence concerning this article should be addressed to Rachel H. Farr, Department of Psychology, University of Massachusetts Amherst, 135 Hicks Way, Amherst, MA 01003. Electronic mail may be sent to rfarr@psych.umass.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0009
DOI: 10.1111/cdev.12046

has remained relatively unexamined. Thus, this study assessed coparenting and associations with child outcomes using a multimethod design among lesbian, gay, and heterosexual couples and their adopted children. As no single study on coparenting has included these three groups, the coparenting literature is reviewed in terms of research with heterosexual couples and biological children, lesbian and gay couples, and adoptive couples.

Coparenting and Child Adjustment

Studying the impact of coparenting on young children's behavioral adjustment is of particular interest. Externalizing behaviors such as aggression, noncompliance, inattention, and hyperactivity often begin during the preschool years and are relatively stable throughout childhood (e.g., Denham et al., 2000; Doctoroff & Arnold, 2004). Efforts to identify children at risk for behavioral and emotional difficulties are often made during early childhood (Phillips & Lonigan, 2010). Furthermore, coparenting has been established as a predictor of children's behavioral adjustment from preschool to school age (Teubert & Pinquart, 2010). For example, Schoppe et al. (2001) found that supportive and undermining coparenting when children were 3 years old was predictive of 4-year-old children's externalizing behaviors. Thus, understanding factors within families that may ameliorate the emergence and developmental course of problem behaviors in preschool-age children is important (Denham et al., 2000), especially among families underrepresented in the literature on coparenting and child outcomes.

Coparenting Among Lesbian and Gay Couples

Coparenting among lesbian and gay couples has generally been studied only in terms of couples' divisions of family labor (Patterson & Farr, 2011), which researchers view as one aspect of coparenting (e.g., Feinberg, 2003). Existing research has shown that lesbian and gay couples often report dividing child-care labor relatively evenly, whereas heterosexual couples often report specialization (Goldberg, 2010). In addition, lesbian parents tend to report ideally wanting an equal distribution of child care between partners. In contrast, heterosexual mothers report ideally wanting to do somewhat more than half of the child care, and heterosexual fathers report ideally wanting to do somewhat less than half (Patterson, Sutfin, & Fulcher, 2004). Furthermore, less research about division of labor has focused on gay fathers than lesbian mothers, and few studies have examined division of child care among adoptive couples (e.g., Bennett, 2003; Goldberg, 2010). Division of labor may be of particular interest among adoptive parents because biological parent–child relationships do not affect family labor decisions (e.g., mothers doing more child care across the transition to parenthood due to pregnancy or breastfeeding; Ward, 1998). Thus, research examining division of labor among a sample of parents where biological relatedness is not a confounding variable is of special interest.

Coparenting, Division of Labor, and Child Adjustment

The extent to which parental division of labor may be associated with children's development, however, remains unclear. Patterson (1995) explored division of labor among lesbian couples and adjustment among young children. When child care was evenly divided, lesbian mothers reported greater satisfaction with divisions of labor and fewer child behavior problems—raising the possibility that shared divisions of labor might, in themselves, be beneficial for children. It is worth noting, however, that since mothers reported that they shared and also that this was their ideal pattern, it was unclear which was more important, or whether other variables might be involved, such as couple relationship satisfaction, role of parental expectations, or parental education. Chan, Brooks, Raboy, and Patterson (1998) also examined division of labor among lesbian and heterosexual couples (who had used donor insemination) and young children's adjustment. Among lesbian nonbiological mothers, those who reported greater satisfaction with division of labor also reported greater couple relationship satisfaction and fewer child behavior problems. The effect of division of labor on children's adjustment was mediated by couple relationship satisfaction. Overall, then, existing research suggests it is parents' feelings about their arrangements rather than actual divisions of labor that are most closely associated with child outcomes (Patterson & Farr, 2011). Further research is needed, however, to clarify these associations, and to examine pathways through which they occur.

Coparenting and Parental Gender

Moreover, there are other important coparenting dynamics beyond division of labor that have not yet been studied in lesbian- and gay-parent families (Patterson & Farr, 2011). Among heterosexual-parent families, it has been debated whether gender-based differences in coparenting by mothers and fathers are relevant to differentiated child outcomes (Brown, Schoppe-Sullivan, Mangelsdorf, & Neff, 2010; Solmeyer & Feinberg, 2011). This question of whether differences in parenting exist on the basis of parental gender has been explored less among lesbian and gay couples, but this research would be informative in addressing the role of parental gender in child development (e.g., Biblarz & Stacey, 2010). While some observational research has included same-sex couples (e.g., Gottman et al., 2003; Julien, Chartrand, Simard, Bouthillier, & Bégin, 2003; Roisman, Clausell, Holland, Fortuna, & Elieff, 2008), no observational data about the relational and emotional dynamics among lesbian- or gay-parenting couples have been reported to date. Thus, questions remain about how coparenting functions among lesbian and gay couples, whether coparenting dynamics among lesbian and gay couples are similar to those among heterosexual couples, and about how coparenting is related to child outcomes in lesbian- and gay-parent families. Specifically, by comparing coparenting behaviors and their associations with child development among lesbian, gay, and heterosexual parent families, the impact of parental gender, as well as differences in the gender composition of couples (i.e., same-sex female, same-sex male, and other-sex male and female) on key family processes can be assessed.

Coparenting Among Adoptive Couples

Few studies of coparenting include adoptive families (e.g., Hock & Mooradian, 2012). Some work has examined relationship dynamics of heterosexual adoptive couples (e.g., Brodzinsky & Huffman, 1988), but not in comparison with lesbian and gay adoptive couples. A few studies have investigated the impact of parental characteristics and parenting style on adoption outcomes (e.g., Palacios, Sanchez-Sandoval, & Leon, 2005). Little research, however, has addressed the associations between coparenting and adopted children's development (Patterson & Farr, 2011), despite a predominant focus in the adoptive families literature on the behavioral and psychological adjustment of adopted children (e.g., Palacios & Brodzinsky, 2010). As adopted children demonstrate greater behavioral difficulties on average than nonadopted children, and because adoptees are overrepresented in clinical settings (e.g., Palacios & Brodzinsky, 2010), careful examination of coparenting relationships, family interactions, and developmental processes among adoptive families could contribute to better understanding of outcomes for adopted children. Adoptive families are a diverse group, including substantial numbers of lesbian and gay parents. The overall numbers of adoptive families have increased in recent years, and the number of lesbian and gay adoptive parents has doubled in the last decade (Gates, 2011). Thus, it is critical for researchers to address the diversity of adoptive families.

The Present Study

Using multiple methods of assessment, we examined coparenting among lesbian, gay, and heterosexual couples with young adopted children. To accomplish three main aims, we studied observations of coparenting, reports of couples' division of child-care labor, perceptions of parent competence, and child outcomes. First, we compared reports about division of child-care labor among lesbian, gay, and heterosexual couples, including satisfaction with child-care arrangements and perceptions of parenting competence. Second, using observational data on family play interactions, we assessed similarities and differences in coparenting among lesbian, gay, and heterosexual couples. Lastly, we examined associations among observations of coparenting, division of child-care labor, parenting competence, and child adjustment across family types.

We had three main hypotheses. Our first hypothesis was that lesbian and gay adoptive parenting couples would be likely to report sharing of child-care tasks, whereas heterosexual adoptive parenting couples would be likely to report specializing, with mothers doing more child care than fathers. We expected observations of coparenting to reveal a similar pattern, with relatively equal involvement among lesbian and gay parents and greater discrepancies among heterosexual parents. Furthermore, on the basis of previous research we expected couples to be relatively satisfied in their divisions of labor, regardless of the way labor was divided. Our second hypothesis was that couples would demonstrate more supportive than undermining behaviors, on average, but that lesbian and gay couples would demonstrate higher levels of cooperation than heterosexual couples. Similarities, rather than differences, were expected in other observations of coparenting. Our third hypothesis was that greater satisfaction with division of child-care labor would be associated with more supportive coparenting and less undermining coparenting, regardless of family type. We also expected that satisfaction with division of child-care labor, rather than actual divisions of labor, would be significantly associated with children's adjustment. Thus, greater parental satisfaction, as well as more supportive and less undermining coparenting were expected to be associated with positive child behavior.

Method

Participants

Data were collected as part of a larger project about child development, parenting, and family functioning in adoptive families with lesbian, gay, and heterosexual parents (e.g., Farr, Forssell, & Patterson, 2010). Participants were recruited through five cooperating adoption agencies in the United States, chosen based on several factors. Agencies had to have placed infants (through domestic adoption) with openly lesbian and gay couples, and agencies had to be situated in jurisdictions permitting same-sex couples to complete legal adoptions. All five agencies also provided options for openness in adoption (i.e., information sharing or contact between birth and adoptive families).

Families in which both parents were legally recognized and currently living with their adopted child between 1 and 5 years old were considered eligible. All were invited via letter or e-mail from the agency director to take part in a study about "child development, parenting, and family relationships in adoptive families." A researcher made follow-up phone calls requesting participation. Recruitment procedures were identical for all families. There were 117 eligible families (23 lesbian, 21 gay, 73 heterosexual couples) from the primary cooperating agency in the Mid-Atlantic United States. Of these, 63 (16 lesbian, 17 gay, and 30 heterosexual) couples participated. Heterosexual couples (41%) were less likely than lesbian and gay couples (75% response rate) to participate, $\chi^2(2, n = 117) = 12.72, p < .01$. Lack of time was the most common reason for nonparticipation. An additional 43 families (11 lesbian, 12 gay, and 20 heterosexual couples) were recruited from four other cooperating agencies in the Northeast, South, and along the West Coast of the United States. Participants contacted the researcher directly to participate after receiving a letter or e-mail of invitation from the agency director. The number of eligible families could not be disclosed due to confidentiality concerns by these agencies; participation rates for this subsample could not be calculated. Two families in the initial sample did not provide sufficient observational data, so they were not included here. Thus, the final sample consisted of 104 families (208 parents, 104 children).

Demographic characteristics of the participating families (54 same-sex couples, 25 lesbian and 29 gay couples, and 50 heterosexual couples) are shown in Table 1. Parents' ages ranged from 30 to 60 years ($M = 42.25$, $SD = 5.83$). Children's ages ranged from 13 to 72 months ($M = 36.07$, $SD = 15.53$). Eighty-one percent of parents were White, 16% were Black or African American, and the remaining 3% were Asian, Latino, or multiracial. Most parents were well educated, worked full-time, and had family incomes above national averages. The sample included 14% interracial couples. Fifty-five families resided in Maryland or the District of Columbia, and others lived in 10 states along the East and West Coasts, or in the Southern United States.

Most families had one child living at home (see Table 1). All parents were the legal parents of their children, adopted domestically. Children (53 girls, 51 boys) had been placed as infants, at birth or within the first few weeks of life. The majority were reported to be healthy, with no special needs. Children were 42% White, 33% Black or African American, 21% multiracial, and 4% other. In the sample, 48% of families had adopted across race (i.e., transracial adoption). About one third of families had some type of direct contact or visitation with birth families a few times per year. All families were English speaking. Parents in seven families spoke an additional language at home (e.g., French, Spanish). The demographic characteristics of this sample closely resembled those of other adoptive families, particularly those completing private domestic infant adoptions (e.g., Brodzinsky & Pinderhughes, 2002).

There were several demographic differences among families. Lesbian mothers had more daughters, whereas gay fathers had more sons. Lesbian and heterosexual couples were less likely to be interracial than gay couples. Same-sex couples completed more transracial adoptions than did heterosexual couples. More details regarding demographic similarities and differences as a function of family type, as well as transracial adoption, are reported elsewhere (Farr & Patterson, 2009). Preliminary analyses revealed no significant associations with reports of division of labor, child adjustment, or coparenting observations; thus, child gender, interracial couple status, and transracial adoptive

Table 1
Demographic Information About Families Headed by Lesbian, Gay, and Heterosexual Parents

	Lesbian mothers (n = 50)	Gay fathers (n = 58)	Heterosexual parents (n = 100)	ANOVA or χ^2 test
Parents (n = 208)				$F(1, 206)$
Mean age at visit (in years)	43 (5)	41 (5)	42 (6)	$F = 2.78$
Race (% White)	80%	86%	78%	$\chi^2 < 1$
Education (% college degree)	94%	89%	85%	$\chi^2 < 1$
Work status (% full-time)	72%	81%	77%	$\chi^2 < 1$
Annual family income ($K)	168 (77)	190 (130)	150 (89)	$F = 1.34$
Interracial relationship	11%	28%	8%	$\chi^2 = 12.14$**
One child in household	63%	62%	50%	$\chi^2 = 1.69$
Transracial adoption	48%	55%	30%	$\chi^2 = 13.27$***
Children (n = 104)				$F(1, 102)$
Mean age at visit (in months)	35 (20)	35 (13)	36 (16)	$F < 1$
Sex (% girls)	59%	36%	52%	$\chi^2 = 3.07$*
Race (% White)	41%	38%	44%	$\chi^2 < 1$
Developmental status (% special needs adoption)	11%	7%	4%	$\chi^2 = 1.44$
Any visits with birthparents? (% yes)	28%	24%	28%	$\chi^2 = 2.79$

Note. Standard deviations are given in parentheses. ANOVA = analysis of variance.
*$p < .05$. **$p < .01$. ***$p < .001$.

status were not included as covariates in further analyses.

There were also 74 teachers or caregivers of the children who provided data (71% response rate). All parents reported that their child was given outside care on a regular basis by some individual, such as a teacher, day-care provider, babysitter, or other relative or adult. Teachers or caregivers answered a few demographic questions and reported on children's behavioral adjustment. Response rates for teachers or caregivers of children with lesbian or gay (n = 42) and heterosexual parents (n = 32) were not different, $\chi^2(1, n = 76) = 2.40$, *ns*. Parents' reports of child behavior did not differ as a function of whether teachers provided data.

Most teachers or caregivers (94%) were female and had attended at least some college (90%). Their average length of experience in teaching or child care was 11.07 years (SD = 8.53). Thus, outside caregivers were generally women with ample experience in their teaching or caregiving roles. Most were preschool (n = 22; 30%) or elementary school (n = 4; 5%) teachers, or day-care center (n = 12; 16%) or home day-care (n = 20; 27%) providers. The other 22% were 11 babysitters and 5 relatives who regularly cared for the child. No significant demographic differences were found among outside caregivers as a function of parental sexual orientation.

Materials and Procedure

A researcher visited families in their homes. Parents responded to paper questionnaires, and families participated in a videotaped, unstructured play session. Families also asked their child's teacher or outside care provider to complete reports of children's behavior problems, mailed back to the researcher in a self-addressed, stamped envelope.

Division of child-care labor. The Who Does What? (Cowan & Cowan, 1990) is designed to assess couples' reports about division of household, decision making, and child-care labor, and is appropriate for use with couples with young children. Child care was the focus here, including items such as feeding, dressing, bathing, choosing toys, and visiting playgrounds. Several minor wording adjustments were made for use with lesbian, gay, and heterosexual couples. On a scale ranging from 1 = *I do it all* to 9 = *my partner or spouse does it all*, parents reported the relative frequency with which they perform each of 20 child-care tasks ("real" involvement scale) or would ideally perform ("ideal" involvement scale). A score of 5 indicated "we do it equally" or "ideally, we would do it equally." Responses were averaged to create one "real" and "ideal" score for child-care tasks. The discrepancy between "real" and "ideal" scores represented the degree to which participants' were satisfied with their current division of labor arrangements. Higher numbers represent greater dissatisfaction with division of labor for child care within the couple. The scales included in the Who Does What? demonstrated high levels of reliability. Both Cronbach's alpha and Spearman-Brown's split-half reliabilities for all subscales ranged from .92 to .99 (Cowan & Cowan, 1990).

One issue in using the Who Does What? with same-sex couples is that there is no simple way to distinguish between the two parents in the couple. With heterosexual couples, parent gender is used to distinguish between mothers and fathers. To address this issue, we created a continuous variable representing specialization. This was calculated by taking the absolute value of the difference between each of the two parents' reports on each of the 20 items, and then calculating the mean for all child-care items. With this new variable, higher values indicated greater specialization of child care within couples, and lower values indicated greater sharing.

Observations of coparenting. As play is an important context for development in early childhood, family play is ideal for observing parent–child interaction, coparenting behavior, and family functioning (Lindsey & Mize, 2001). A number of researchers have developed coding systems of triadic family interaction based on observations of unstructured play in families with children from infancy to preschool; these play sessions were set up in the lab or home with toys provided (e.g., McHale et al., 2001). Thus, in this study, an unstructured family play session was used to observe coparenting interactions. A simple fleece blanket with dimensions of 54 in. × 66 in. (137 cm × 168 cm) was used to designate a play area in participants' homes, and a camcorder was set up on a tripod to record family play. To optimize the appropriateness of the session across the age range of children in this study, toys were provided in one of two sets—one designed for toddlers (1 to 2½ years old) and one for preschoolers (3 to 5 years old). Parents and children were invited to play for 10 min using the provided toys.

The Coparenting Behavior Coding scale (Cowan & Cowan, 1996; Schoppe et al., 2001) was designed to assess coparenting during everyday family interactions in two-parent families with young children. The scale includes supportive and undermining dimensions of coparenting on 5-point rating scales (1 = *very low*, 5 = *very high*). Trained coders were instructed to focus on interactions between parents with reference to the child or to parenting. A Supportive dimension score was created by computing the mean of subscale scores for pleasure, cooperation, interactiveness, and warmth. An Undermining dimension score was created by computing the mean of subscale scores for displeasure, coldness, anger, and competition. Consistent with research on coparenting in earlier studies, Supportive and Undermining dimensions were not correlated, $r(104) = .16$, $p = .11$ (e.g., Schoppe-Sullivan, Mangelsdorf, Frosch, & McHale, 2004). In addition, individual parent participation was rated on a scale of 1 to 7, with higher numbers indicating greater participation in family interaction, adapted from the family problem solving code (Forbes, Vuchinich, & Kneedler, 2001).

Two students served as coders of the family interactions. Coders spent a minimum of 20 hr in training to learn the coding system, spending time viewing, rating, and discussing a set of pilot family play sessions. After achieving high reliability (alpha $\geq .80$), coders independently rated family interactions for 50% of families. Both coders rated an additional 25% of family interactions to establish reliability. For ratings of parent participation, separate coders observed each parent in a family. Coders were closely monitored, with ample opportunities for feedback. There were checkpoints for reliability at the 25%, 50%, and 75% completion marks. In the case of discrepancies, a third trained coder broke the tie. Reliabilities for the eight individual coparenting dimensions as well as parent participation ranged from .84 to .96 ($M = .91$).

Child adjustment. Children's behavioral adjustment was assessed using the Child Behavior Checklist (CBCL) and the Caregiver–Teacher Report Form (C–TRF) for 1½ to 5 year olds (Achenbach & Rescorla, 2000). Both include a total problem score, as well as internalizing and externalizing scores. All 100 behavior problem items are rated by parents and teachers from 0 to 2 (0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very true or often true*). The focus of this study was on the externalizing subscale, composed of 24 CBCL items and 34 C–TRF items assessing children's disruptive, aggressive, and delinquent behaviors (e.g., "Hits others"). Age- and sex-specific raw scores can be converted to standard *T* scores. Higher scores suggest greater behavior problems. The CBCL and C–TRF are widely used and have demonstrated high levels of reliability and validity (Achenbach & Rescorla, 2000). Cronbach's alphas for externalizing scores for the sample were .85 for the CBCL and .92 for the C–TRF. Alphas for CBCL and C–TRF externalizing scores were .89 and .95 in lesbian mother families, .90

and .87 in gay father families, and .81 and .90 in heterosexual parent families, respectively.

Preliminary Analyses

Power analyses (alpha = .05) were conducted to determine power levels for the analyses of interest ($N = 104$ families). For bivariate correlations among all families, power reached .99 for large and medium effects, and .88 for small effects. For chi-square tests with 2 *df* among three family groups, power reached .99 for large, .79 for medium, and .13 for small effects. For multivariate analyses of variance (MANOVAs) measuring global effects with two response variables (supportive and undermining coparenting) among three family groups, power reached .99 for large, .99 for medium, and .97 for small effects. For analyses of variance (ANOVAs) of main effects and interactions with three groups, power reached .96 for large, .61 for medium, and .14 for small effects. For multiple regression with three predictors, power reached .99 for large, .92 for medium, and .19 for small effects. We conclude that although not all of our analyses were sufficiently powered to detect small effects (e.g., $d = .20$, $f^2 = .10$), they were sufficiently powered to detect medium effects ($d = .50$, $f^2 = .25$) and more than adequately powered to detect large effects ($d = .80$, $f^2 = .40$).

To evaluate the role of children's age and gender, initial analyses explored associations among child age, gender, and variables of interest. Boys scored higher than girls on externalizing problems as reported by parents ($M_{\text{boys}} = 47.68$, $SD = 9.11$; $M_{\text{girls}} = 45.20$, $SD = 8.74$), $t(103) = 2.00$, $p = .047$. Teacher reports showed no significant differences in externalizing problems as a function of child gender. No coparenting observations or parent-report measures yielded significant associations with child age or gender. There were no interactions of child age or gender with variables of interest, including family type. Thus, child age and gender were excluded from further analysis.

Data Analytic Plan

Hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) was used to account for the nested structure of the data. The two parents in each family were not independent of one another in their individual reports of satisfaction with division of labor. In statistical terms, parents were nested within families. Thus, HLM was used to control for sources of shared variance and data dependency within families. One challenge that arises in using HLM among a sample of same-sex and other-sex couples is that the models must account for both indistinguishable dyads (i.e., lesbian and gay couples) and distinguishable dyads (i.e., heterosexual couples; e.g., Goldberg, Smith, & Kashy, 2010). To examine hypotheses regarding associations of family-type and couple-level variables (i.e., satisfaction with division of labor), we followed the methods of previous researchers working with indistinguishable and distinguishable dyads, particularly lesbian, gay, and heterosexual couples (Goldberg et al., 2010; Kurdek, 1998).

In analyses comparing by couple type, the conditional models included individual partners (Level 1) nested in couples (Level 2). The intercept at Level 1 represented the outcome variable explained at Level 2. The within-couples model at Level 1 used information from both partners to define one parameter (the intercept) for each couple. As Kurdek (1998) described, this intercept is regarded as a random variable because the couple-level intercepts obtained were derived from a larger population of couple-level intercepts. Level 2 represented the between-couples model comparing the effect of being "lesbian versus heterosexual" and "gay versus heterosexual" for the outcome variables included (i.e., satisfaction with division of child care and perceptions of parenting competence). The Level 2 intercept thus corresponded to the mean ratings for heterosexual parents. This conditional model can be described in Level 1: $Y_{ij} = \beta_{0j} + e_{ij}$ and Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Lesbian}) + \gamma_{02}(\text{Gay}) + u_{0j}$. In the Level 1 equation, the outcome variable was Y_{ij}. The random intercepts were represented by β_{0j} coefficient. The error term was e_{ij}. Level 1 reflected the average calculated for each outcome variable. At Level 2, the γ_{01}(Lesbian) coefficient represented the "lesbian versus heterosexual effect" whereas γ_{02}(Gay) represented the "gay versus heterosexual effect." The u_{0j} coefficient controlled for the dependency of partners' data within couples. Level 2 reflected a comparison of averages for each of the outcome variables to examine differences by family type.

Results

Primary Analyses

Results are presented in three main sections: (a) parent reports of division of child-care labor, (b) similarities and differences in coparenting observations among family types, and (c) associations among division of labor, observations of coparenting, and child adjustment.

Division of Child-Care Labor

As expected, lesbian and gay parents were more likely than heterosexual parents to report sharing child-care labor (see Table 2). Specialization was significantly greater among heterosexual couples than among lesbian and gay couples. On average, mothers reported doing significantly more child care than fathers in heterosexual couples. On average, parents reported being relatively satisfied with their current divisions of child-care labor ($M = 0.45$, $SD = 0.46$) with no differences by family type (see Table 2). Among heterosexual couples, however, mothers were significantly less satisfied with current child-care arrangements than fathers. HLM results were consistent with the ANOVA results regarding satisfaction of child-care labor divisions.

In sum, reports about division of labor varied across and within couples. Consistent with our expectations, lesbian and gay couples were more likely to report sharing child care, and heterosexual couples were more likely to report specialization. Regardless of their actual arrangements, most parents reported feeling satisfied with them.

Observations of Coparenting Behavior

On average, family interactions were characterized by higher levels of supportive coparenting practices ($M = 2.84$, $SD = 0.59$) than undermining practices ($M = 1.55$, $SD = 0.44$). However, as expected, observations of coparenting varied by family type (see Table 3). MANOVAs revealed significant differences as a function of family type in both the composite scores for supportive and undermining coparenting dimensions. Lesbian couples showed the most supportive interactions, whereas gay couples showed the least. Heterosexual couples were intermediate between lesbian and gay couples in supportive behaviors. Regarding undermining interactions, lesbian couples were the least, and heterosexual couples were the most undermining. Gay couples were intermediate between lesbian and heterosexual couples.

Follow-up ANOVAs were conducted to explore group differences among individual coparenting dimensions (see Table 3), particularly because this was the first time this coding scheme had been utilized with same-sex couples and adoptive couples (S. J. Schoppe-Sullivan, personal communication, March 2011). There were significant differences in pleasure, warmth, and interactiveness by family type. Lesbian couples showed the most pleasure of interaction, gay couples the least, and heterosexual couples were intermediate between the two other couple types. Lesbian couples were also the most warm, gay couples the least, and heterosexual couples were intermediate. Interactiveness followed this pattern, with lesbian couples demonstrating the most interaction, gay couples the least, and heterosexual couples intermediate between the two. Cooperation between parents did not, however,

Table 2
Division of Child-Care Labor Reported Among Lesbian, Gay, and Heterosexual Couples

	Lesbian mothers ($n = 50$)	Gay fathers ($n = 58$)	Heterosexual parents ($n = 100$)	ANOVA $F(5, 202)$
"How it is now"[a]				
Parent A	5.36 (0.95)	5.38 (0.68)	6.11 (0.85)$_a$	
Parent B	4.80 (0.88)	4.97 (0.70)	4.13 (0.74)$_b$	32.57***
				$F(2, 101)$
Specialization	1.97 (1.05)$_a$	1.88 (0.87)$_a$	2.59 (1.14)$_b$	5.27**
Dissatisfaction with arrangement[a]				$F(5, 202)$
Parent A	0.51 (0.52)	0.42 (0.49)	0.66 (0.57)$_a$	
Parent B	0.38 (0.45)	0.38 (0.31)	0.32 (0.28)$_b$	3.25**

Note. Standard deviations are given in parentheses. Means in a row with different subscripts were found to be significantly different using Tukey post hoc tests. Scores for "How it is now" range from 1 to 9, with 1 = *my partner does it all* and 9 = *I do it all*. Higher scores for Specialization reflect a greater degree of specialization in child-care labor within a couple (a score of 0 would reflect a perfectly shared division of child-care labor). Dissatisfaction scores represent a discrepancy between parents' reports of "real" and "ideal" division of child-care arrangements. Higher scores reflect greater dissatisfaction.

[a] In heterosexual couples, Parent A was always the mother and Parent B was always the father. In same-sex couples, Parent A was assigned to the first person who responded to participate in the study. Analysis of variance (ANOVA) compares the six groups of parents in this case (Parent A and Parent B in each of the three family types).

$^{**}p < .01$. $^{***}p < .001$.

Table 3
Observations of Coparenting Among Lesbian, Gay, and Heterosexual Parenting Couples

	Lesbian couples (n = 25)	Gay couples (n = 29)	Heterosexual couples (n = 50)	Average (N = 104)	ANOVA F(2, 101)
Supportive	3.12 (0.61)$_a$	2.56 (0.48)$_b$	2.87 (0.56)$_a$	2.84 (0.59)	6.93**
Pleasure	3.08 (0.70)$_a$	2.55 (0.69)$_b$	2.72 (0.67)	2.76 (0.70)	4.18*
Warmth	3.12 (0.88)$_a$	2.31 (0.66)$_b$	2.64 (0.69)$_b$	2.66 (0.78)	8.22***
Interactiveness	3.20 (0.87)$_a$	2.48 (0.83)$_b$	3.02 (0.89)$_a$	2.91 (0.90)	5.30**
Cooperation	3.08 (0.70)	2.90 (0.62)	3.10 (0.65)	3.04 (0.65)	< 1 *ns*
Undermining	1.38 (0.38)$_a$	1.49 (0.37)	1.67 (0.46)$_b$	1.55 (0.44)	4.33*
Displeasure	1.04 (0.20)$_a$	1.24 (0.64)$_a$	1.60 (0.76)$_b$	1.37 (0.67)	7.33**
Coldness	1.44 (0.65)	1.38 (0.56)	1.64 (0.69)	1.52 (0.65)	< 1 *ns*
Anger	1.04 (0.20)	1.07 (0.37)	1.24 (0.43)	1.14 (0.38)	3.34*
Competition	2.00 (0.76)	2.28 (0.84)	2.20 (0.78)	2.17 (0.79)	< 1 *ns*
Parent participation[a]	(n = 50)	(n = 58)	(n = 100)	(N = 208)	F(5, 202)
Parent A	4.88 (1.01)$_{a,c}$	4.28 (1.00)$_{d,e}$	4.98 (0.96)$_{a,c}$		
Parent B	5.12 (0.97)$_{a,c}$	4.48 (0.95)	4.04 (1.23)$_{b,d}$	4.59 (1.11)	6.46***
					F(2, 101)
Discrepancy in participation	1.04 (0.83)$_a$	1.10 (0.85)$_a$	1.46 (0.95)$_b$	1.26 (0.91)	4.90**

Note. Standard deviations are given in parentheses. Means in a row with different subscripts were found to be significantly different using Tukey post hoc tests.
[a] In heterosexual couples, Parent A was always the mother and Parent B was always the father. In same-sex couples, Parent A was assigned to the first person who responded to participate in the study. Analysis of variance (ANOVA) compares the six groups of parents in this case (Parent A and Parent B in each of the three family types).
*$p < .05$. **$p < .01$. ***$p < .001$.

differ among family types, with an average score of 3.04 (SD = 0.65) across families. Family type was also a predictor of displeasure and anger. Lesbian couples showed the least displeasure, heterosexual couples the most, and gay couples were intermediate. Heterosexual couples showed significantly more anger during interaction than did lesbian or gay couples. Coldness and competition did not vary by family type.

Overall, and consistent with our expectations, parents showed moderately high levels of involvement during family play (M = 4.59, SD = 1.11). There were, however, significant differences among family groups (see Table 3). Post hoc Tukey analysis revealed that, as expected, heterosexual fathers participated significantly less ($M_{\text{Het. Fathers}} = 4.04$, SD = 1.23) than did heterosexual mothers ($M_{\text{Het. Mothers}} = 4.98$, SD = 0.96) or lesbian mothers ($M_{\text{Lesbian}} = 5.00$, SD = 0.99), $F(5, 202) = 6.46$, $p < .001$. In examining discrepancies in participation between parents in each couple, heterosexual couples had significantly greater discrepancies in participation than did lesbian and gay couples. Overall, fathers participated less during family interaction than mothers in heterosexual couples, consistent with parents' reports of divisions of child-care labor.

Associations Among Coparenting Observations, Division of Labor, and Child Adjustment

Parents and teachers described children as showing relatively few externalizing problems on average (CBCL: M = 46.41, SD = 8.99; C-TRF: M = 50.08, SD = 8.63), with no differences as a function of family type. For more details about children's behavioral adjustment and family structure using HLM in this sample, please see Farr et al. (2010). Associations among all variables related to coparenting observations, divisions of child-care labor, and child externalizing problems as reported by parents and teachers are presented in Table 4.

Data reduction. Data reduction was undertaken, similar to the methods of Schoppe et al. (2001), to facilitate comparisons of children's adjustment with coparenting observations and division of labor reports. Parents' scores within couples for child externalizing problems were significantly correlated, $r(208) = .48$, $p < .001$ (correlations were significant for all family types, ranging from .38 to .66), so the two-parent reports for each family were summed in a composite score. Teachers' and parents' reports were significantly associated, $r(74) = .19$, $p = .023$,

Table 4

Correlations Among Observations of Coparenting, Division of Labor, and Child Behavioral Adjustment

Variable	1	a	b	c	d	2	e	f	g	h	3	4	5	6	7	8
1. Supportive	—															
a. Pleasure	.81***	—														
b. Warmth	.89***	.73***	—													
c. Interaction	.77***	.44***	.58***	—												
d. Cooperation	.61***	.35***	.40***	.24*	—											
2. Undermining	−.14	−.07	−.16	−.01	−.23*	—										
e. Displeasure	.12	.12	.07	.14	.01	.73***	—									
f. Coldness	−.29**	−.25**	−.29**	−.14	−.24*	.65***	.20*	—								
g. Anger	−.04	−.03	−.04	.17†	−.28**	.62***	.58***	.17†	—							
h. Competition	−.14	−.03	−.15	−.10	−.19†	.76***	.31**	.36***	.26**	—						
3. Dissatisfaction (DoL; mean)^{a}	.07	.18†	.08	−.03	.00	−.01	.14	−.07	.01	−.10	—					
4. Specialization	.08	.08	.03	.04	.09	.06	.18†	.02	.04	−.06	.54***	—				
5. Participation (mean)^{a}	.54***	.46***	.45***	.34**	.44***	−.01	.07	−.21*	.08	.04	.03	.08	—			
6. Participation discrepancy	−.16	−.05	−.14	−.10	−.20*	−.17†	−.04	.22*	.10	.18†	.15	.12	−.23*	—		
7. CBCL Ext^{a}	−.22*	−.19*	−.15	−.24*	−.06	.15	.10	.02	.06	.20*	.20*	.14	−.08	.15	—	
8. C–TRF Ext	−.03	−.09	.09	−.04	−.07	−.10	−.16	−.17	.06	.00	−.16	−.16	−.19	.15	.22†	—

Note. Pearson product moment correlations calculated for all variables. CBCL Ext represents parents' reports ($N = 104$ families) of child externalizing behavior problems, whereas C–TRF Ext represents teachers' reports ($N = 74$) of child externalizing behavior problems.
^{a}Mean scores reflect the average scores of the two parents within each couple.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

but only moderately, so teachers' and parents' scores for child externalizing problems were examined separately in further analyses.

No significant association was found between children's adjustment and the degree to which couples specialized in child care. However, consistent with earlier findings, children's externalizing problems (as reported by parents) were significantly associated with couples' reported satisfaction with child-care arrangements, $r(104) = .20$, $p = .043$; parents who reported less satisfaction described their children as having more externalizing behaviors. Contrary to expectations, coparenting observations and parents' satisfaction with divisions of labor were not significantly associated.

Some observations of coparenting among lesbian, gay, and heterosexual parents were associated with parent and teacher reports of children's behavioral adjustment. The composite supportive coparenting score was associated with parents' reports of child externalizing problems, $r(104) = -.21$, $p = .034$. Fewer externalizing problems, as reported by parents, were also associated with two subscales of supportive coparenting: pleasure of interaction, $r(104) = -.20$, $p = .044$, and greater interaction between parents, $r(104) = -.23$, $p = .020$. Undermining coparenting behavior was significantly associated with parent reports of child behavior problems such that greater competition was associated with greater externalizing problems, $r(104) = .22$, $p = .027$. Teachers' reports of child externalizing problems were, however, not associated with coparenting observations. Overall, it appeared that aspects of both supportive and undermining coparenting were relevant to children's externalizing behaviors, as reported by parents.

Next, we conducted hierarchical multiple regression analyses predicting child externalizing behaviors from coparenting observations and division of labor variables that were significantly associated with parent reports of children's behavioral adjustment (see Table 5). Simple linear regressions revealed that fewer child externalizing problems were significantly predicted by observations of coparenting: greater supportive coparenting, $F(1, 102) = 4.61$, $p = .034$, greater interactiveness; $F(1, 102) = 5.61$, $p = .020$; greater pleasure of interaction, $F(1, 102) = 4.15$, $p = .044$; and less competition between parents, $F(1, 102) = 5.04$, $p = .027$. Fewer child externalizing problems, as reported by parents, were also predicted by greater satisfaction

Table 5

Simultaneous Multiple Linear Regression of Coparenting and Division of Labor Variables Onto Summed Parent Reports of Child Externalizing Problems

	Child externalizing problems			
Variable	*B*	*SE* (*B*)	β	*t*
Supportive	−2.85	1.47	−.18	$-1.94^{\dagger}$
Competition	3.11	1.47	.20	2.11*
Dissatisfaction (child-care divisions)	3.48	1.45	.23	2.40*
R^2	.13			
F(3, 100)	4.84**			

Note. All predictor variables were centered. Satisfaction with child-care divisions represents an average satisfaction score for each couple. $^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

with division of child-care labor, $F(1, 102) = 4.20$, $p = .043$.

To examine whether these associations differed as a function of family type, we recalculated all simple linear regressions that were significant in the previous step with family type as a moderator. Family type was dummy coded to evaluate "lesbian versus heterosexual" and "gay versus heterosexual" effects with heterosexual parents as the reference group. All independent variables were centered. In no case was the moderation by family type significant, indicating that the associations among observations of coparenting, division of labor variables, and child externalizing problems did not differ by parental sexual orientation.

Next, the variables that were significant in predicting child externalizing problems in the simple linear regressions were entered into a multiple regression model. Dissatisfaction with child-care division of labor, as well as competition between parents, emerged as significant (see Table 5). The model explained 13% of the variance in child externalizing problems.

Thus, as expected, parents' satisfaction with division of child-care labor, not their actual division of labor, was significantly associated with children's behavior problems. As anticipated, supportive coparenting interactions, such as those characterized by greater pleasure and engagement between parents, were associated with positive child behavior. Undermining coparenting, particularly competition, was associated with child conduct problems. These results held for all three types of families. Parental division of labor as such (whether assessed via parental reports or behavioral observations) was not, however, associated with child outcomes.

Discussion

The current study provided the first empirical examination of differences and similarities in coparenting among lesbian, gay, and heterosexual couples, as well as an evaluation of their associations with outcomes among adopted children. Results showed that lesbian and gay couples were more likely than heterosexual couples to share parenting tasks evenly, and that heterosexual couples were more likely to specialize. Whether parents shared child care or showed a more specialized pattern was, however, unrelated to children's adjustment. The best predictors of child behavior problems were observed competition between parents, and dissatisfaction with child-care labor divisions, which were not associated with parental sexual orientation. Our results thus revealed both similarities and differences among the coparenting interactions of lesbian, gay, and heterosexual couples, and they suggested that group differences in coparenting as such were unrelated to child adjustment.

As expected, both lesbian and gay couples were likely to report sharing child-care tasks and heterosexual couples were likely to report specialization. These findings expand on earlier findings that lesbian parents are more likely to report sharing labor than are heterosexual parents (Goldberg, 2010). These data are among the first to reveal that gay fathers divide child-care tasks relatively evenly (see also Johnson & O'Connor, 2002).

Despite differences in their reported patterns of dividing child-care labor, most parents reported feeling equally satisfied with their current divisions of labor. Couples were more likely to express dissatisfaction with divisions of labor when they reported greater specialization—suggesting a possible disadvantage of dividing labor according to traditional gender roles. Mothers in heterosexual couples, who reported doing more child care than fathers, were also more likely than fathers to report dissatisfaction with current child-care arrangements. Similarly, previous research indicates that inequities in divisions of child-care labor are often associated with dissatisfaction and lower relationship quality, particularly among heterosexual women (e.g., Coltrane, 2000). Thus, advantages of sharing child-care tasks may stem from greater satisfaction with division of labor. Indeed, more egalitarian (less "traditional") divisions of labor in heterosexual couples have been associated with higher marital satisfaction (e.g., Coltrane, 2000).

Consistent with coparenting observations of heterosexual couples with young biological children

(Jia & Schoppe-Sullivan, 2011), adoptive couples demonstrated relatively high levels of supportive coparenting and low levels of undermining coparenting behaviors during family play. Supportive behaviors were observed in the degree of warmth, enjoyment of interaction, and level of interactiveness between parents. An example of warmth was expressed during a family's "tea party" when one gay father handed the other a tea cup, and the partners smiled at one another, maintained eye contact, and laughed together. Interactiveness and enjoyment of interaction were observed with one lesbian couple when they discussed what the child had done earlier that day while smiling and laughing together. Undermining behaviors were also observed. For example, one heterosexual couple, rated high in competition, was observed talking over one another, with each parent suggesting different toys to their child. Across all families, however, supportive behaviors were significantly more likely than undermining behaviors. These results mirror those of other studies using the same observational coding scheme (Jia & Schoppe-Sullivan, 2011) and extend earlier findings for the first time to lesbian, gay, and heterosexual adoptive couples.

Significant differences were, however, found in coparenting such that a unique profile of coparenting dynamics emerged for each family type. Lesbian couples were rated as the most supportive and the least undermining. Gay couples were rated as the least supportive, but were intermediate to lesbian and heterosexual couples in undermining behaviors. Heterosexual couples were intermediate in supportive behaviors, but showed the most undermining behaviors.

These findings about differences in coparenting behavior indicate how parental sexual orientation, as well as parent gender, may be linked with important qualitative differences in the experiences of adoptive families. For example, results showed that the more women in the couple, the more supportive the interactions were rated as being. This finding could be related to the ways in which women are typically socialized to be warm, nurturing, and relational in their interactions with others (Patterson & Hastings, 2007). It also may be that men demonstrate supportive behaviors in ways that differ from the ways that women typically do. We found that heterosexual couples demonstrated significantly more undermining behaviors than did lesbian or gay couples. This could be related to societal power differentials between women and men in heterosexual couples as compared to lesbian and gay couples (Goldberg & Perry-Jenkins, 2007) or to differences in power and influence among heterosexual couples in the coparenting and couple relationship (McHale & Irace, 2011). Alternatively, this result could reflect the greater value that lesbian and gay couples may place on egalitarian relationship behavior as compared to heterosexual couples (Goldberg & Perry-Jenkins, 2007). Overall, these results reflect a complex interplay of gender and sexual orientation among adoptive couples.

Contrary to our hypotheses, cooperation and competition were not observed to differ by family type. One of the reasons for this may be the relative homogeny of this sample, with regard to low ratings of undermining behavior and high ratings of supportive behavior overall. These findings are perhaps not surprising given that these were all adoptive parenting couples. Earlier research clearly demonstrates that adoptive parents have similar, or sometimes more favorable, psychological adjustment than do other parents—for instance, adoptive parents often report less parenting stress than do biological parents (Brodzinsky & Pinderhughes, 2002). This may reflect the fact that adoptive parents complete a rigorous screening process before completing an adoption, in which prospective parents are evaluated particularly in terms of having positive mental health and high couple relationship quality (Brodzinsky & Pinderhughes, 2002). More recently, Hock and Mooradian (2012) note that coparenting among adoptive couples may be more stable than that for other couples. Thus, it is not surprising that most couples conveyed supportive, rather than undermining, coparenting interactions.

All parents, on average, were involved in family interaction. Consistent with division of labor reports, however, heterosexual couples had greater discrepancies in participation than did lesbian and gay couples. Heterosexual mothers were observed to participate more than did heterosexual fathers, in contrast to same-sex couples, who tended to participate more equally. The results support the validity of Cowan and Cowan's (1990) widely used self-report measure.

Parents' subjective evaluations of how child-care labor was divided were significantly associated with children's behavior problems, such that parents who reported greater satisfaction with child-care arrangements also reported that their children were more well adjusted. The degree to which couples specialized in child-care labor, however, was not associated with children's adjustment. Thus, consistent with the findings of Chan et al. (1998), couples' satisfaction with division of labor was more clearly linked with child outcomes than was their actual division of labor.

Consistent with earlier research on heterosexual couples with young biological children (Teubert & Pinquart, 2010), observations of coparenting were significantly related to parents' reports of children's adjustment. These associations did not differ by family type, suggesting that some family processes are related to child outcomes regardless of family structure. In particular, competition between parents was related to more child externalizing problems as reported by parents. It may be that different predictors emerged as significant based from parent reports, but not from teacher reports, due to the different contexts in which parents and teachers observe children's behaviors. Parents may be observing behaviors relevant to parenting and family contexts, whereas teachers may be capturing behaviors related to group and social settings in their ratings (Mangelsdorf, Schoppe, & Buur, 2000).

Contrary to expectations, supportive and undermining coparenting behaviors were not significantly associated with reported divisions of labor or with parents' satisfaction with these arrangements. Family systems researchers have argued that the coparenting relationship should be examined separately from the couple relationship (Schoppe-Sullivan et al., 2004), and these results appear consistent with that view. Although Feinberg (2003) has discussed division of family labor as one aspect of coparenting, he acknowledged that different components of coparenting are likely to be partly associated and partly distinct. In this study, measurements of division of labor may have captured dynamics of the couple's relationship as partners, whereas observations of coparenting may have captured the dynamics of the partners' relationships specific to their parenting roles. This divergence has been observed in families with infants (Van Egeren, 2004), and the current results extend these findings to families with young children. Thus, it appears important to examine division of labor and other coparenting dynamics as separate constructs.

Strengths and Limitations

This study had a number of strengths. It was the first to address coparenting dynamics and children's development in a sample of adoptive families headed by lesbian, gay, and heterosexual parents. These are among the first observational data on coparenting to be collected from lesbian- and gay-parent families, as well as from adoptive families (Hock & Mooradian, 2012). Thus, the findings contribute to the literatures on developmental psychology and family systems, and also to the limited knowledge base regarding gay father families and adoptive families with lesbian and gay parents (Goldberg, 2010). Because aggregating data from multiple informants is likely to yield a more complete picture of children's adjustment (e.g., Doctoroff & Arnold, 2004), child outcomes were assessed not only from parent reports but also from teacher reports. The sample was systematically recruited from a variety of geographical locations and assessed using multiple methods, likely enhancing the capacity for generalizing the findings.

Some limitations of the study should be noted. At the time of data collection, children were young ($M = 3$ years) and thus this study was not likely to have captured complex dynamics of parenting and child development that emerge as children grow older. Longitudinal data would be valuable. Our observational play sessions were limited in length, and additional data about family interaction at multiple time points, across multiple settings, and from multiple informants would be helpful. Families in which both parents had been awarded legal recognition as parents participated, so it remains to be seen whether the findings would hold true in jurisdictions in which only one partner in a same-sex couple is the legal parent. Participation rates among heterosexual couples were lower than those among lesbian and gay couples, making generalization beyond this sample uncertain. Broadly, future research about how contexts within and outside families (e.g., sibling relationships, contact with birth families, transracial adoption, peer relationships, and school experiences) affect child development and family dynamics in adoptive families with lesbian and gay parents would be valuable.

Conclusion

In conclusion, the results revealed both similarities and differences in coparenting styles among lesbian, gay, and heterosexual couples. Although couples differed in their reports of divisions of child-care labor and in observational assessments of coparenting behaviors, their young adopted children were reported to have few behavioral adjustment problems. Overall, coparenting behaviors emerged as more important correlates of child outcomes than did family structure, regardless of differences in couples' divisions of labor or interactions during family play. Thus, results suggested that parental sexual orientation was linked more with qualitative differences in family experiences than with differences in outcomes for children.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms profiles*. Burlington: University of Vermont Research Center for Children, Youth & Families.

Bennett, S. (2003). Is there a primary mom? Parental perceptions of attachment bond hierarchies within lesbian adoptive families. *Social Work, 20*, 159–173.

Biblarz, T. J., & Stacey, J. (2010). How does the gender of parents matter? *Journal of Marriage and Family, 72*, 3–22. doi:10.1111/j.1741-3737.2009.00678.x

Bornstein, M. H., & Sawyer, J. (2006). Family systems. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook on early childhood development* (pp. 381–398). Malden, MA: Blackwell.

Brodzinsky, D. M., & Huffman, L. (1988). Transition to adoptive parenthood. *Clinical Psychology Review, 7*, 25–47. doi:10.1300/J002v12n03_13

Brodzinsky, D. M., & Pinderhughes, E. (2002). Parenting and child development in adoptive families. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 1. Children and parenting* (2nd ed., pp. 279–312). Mahwah, NJ: Erlbaum.

Brown, G. L., Schoppe-Sullivan, S. J., Mangelsdorf, S. C., & Neff, C. (2010). Observed and reported supportive coparenting as predictors of infant-mother and infant-father attachment security. *Early Child Development and Care, 180*, 121–137. doi:10.1080/03004430903415015

Chan, R. W., Brooks, R. C., Raboy, B., & Patterson, C. J. (1998). Division of labor among lesbian and heterosexual parents: Associations with children's adjustment. *Journal of Family Psychology, 12*, 402–419. doi:10.1037/0893-3200.12.3.402

Coltrane, S. (2000). Research on household labor: Modeling and measuring the social embeddedness of routine family work. *Journal of Marriage and Family, 62*, 1208–1233. doi:10.1111/j.1741-3737.2000.01208.x

Cowan, C. P., & Cowan, P. A. (1990). Who does what? In J. Touliatos, B. F. Perlmutter, & M. A. Straus (Eds.), *Handbook of family measurement techniques* (pp. 447–448). Beverly Hills, CA: Sage.

Cowan, C. P., & Cowan, P. A. (1996). *Schoolchildren and their families project: Description of coparenting style ratings*. Unpublished coding scales, University of California, Berkeley.

Denham, S. A., Workman, E., Cole, P. M., Weissbrod, C., Kendziora, K. T., & Zahn-Waxler, C. (2000). Prediction of externalizing behavior problems from early to middle childhood: The role of parental socialization and emotion expression. *Development and Psychopathology, 12*, 23–45. doi:10.1017/S0954579400001024

Doctoroff, G. L., & Arnold, D. H. (2004). Parent-rated externalizing behavior in preschoolers: The predictive utility of structured interviews, teacher reports, and classroom observations. *Journal of Clinical Child and Adolescent Psychology, 33*, 813–818. doi:10.1207/s15374424jccp3304_16

Farr, R. H., Forssell, S. L., & Patterson, C. J. (2010). Parenting and child development in adoptive families: Does parental sexual orientation matter? *Applied Developmental Science, 14*, 164–178. doi:10.1080/10888691.2010.500958

Farr, R. H., & Patterson, C. J. (2009). Transracial adoption by lesbian, gay, and heterosexual parents: Who completes transracial adoptions and with what results? *Adoption Quarterly, 12*, 187–204. doi:10.1080/10926750903313328

Feinberg, M. E. (2003). The internal structure and ecological context of coparenting: A framework for research and intervention. *Parenting: Science and Practice, 3*, 95–131. doi:10.1207/S15327922PAR0302_01

Forbes, C., Vuchinich, S., & Kneedler, B. (2001). Assessing families with the family problem solving code. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 59–76). Mahwah, NJ: Erlbaum.

Gates, G. J. (2011). Family formation and same-sex couples raising children. *National Council of Family Relations Family Focus on LGBT Families Newsletter, FF51*, 1–4.

Goldberg, A. E. (2010). *Lesbian and gay parents and their children: Research on the family life cycle*. Washington, DC: American Psychological Association.

Goldberg, A. E., & Perry-Jenkins, M. (2007). The division of labor and perceptions of parental roles: Lesbian couples across the transition to parenthood. *Journal of Social and Personal Relationships, 24*, 297–318. doi:10.1177/0265407507075415

Goldberg, A. E., Smith, J. Z., & Kashy, D. A. (2010). Preadoptive factors predicting lesbian, gay, and heterosexual couples' relationship quality across the transition to adoptive parenthood. *Journal of Family Psychology, 24*, 221–232. doi:10.1037/a0019615

Gottman, J., Levenson, R., Swanson, C., Swanson, K., Tyson, R., & Yoshimoto, D. (2003). Observing gay, lesbian and heterosexual couples' relationships: Mathematical modeling of conflict interaction. *Journal of Homosexuality, 45*, 65–91. doi:10.1300/J082v45n01_04

Hock, R. M., & Mooradian, J. K. (2012). Co-parenting quality among adoptive mothers: Contributions of socioeconomic status, child demands and adult relationship characteristics. *Child & Family Social Work, 17*, 85–95. doi:10.1111/j.1365-2206.2011.00775.x

Jia, R., & Schoppe-Sullivan, S. J. (2011). Relations between coparenting and father involvement in families with preschool-age children. *Developmental Psychology, 47*, 106–118. doi:10.1037/a0020802

Johnson, S. M., & O'Connor, E. (2002). *The gay baby boom: The psychology of gay parenthood*. New York: New York University Press.

Julien, D., Chartrand, E., Simard, M., Bouthillier, D., & Bégin, J. (2003). Conflict, social support and relationship quality: An observational study of heterosexual, gay male and lesbian couples' communication. *Journal of Family Psychology, 17,* 419–428. doi:10.1037/0893-3200.17.3.419

Kurdek, L. A. (1998). Relationship outcomes and their predictors: Longitudinal evidence from heterosexual married, gay cohabiting, and lesbian cohabiting couples. *Family Relations, 60,* 553–568. doi:10.2307/353528

Lindsey, E. W., & Mize, J. (2001). Measuring parent-child mutuality during play. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 167–182). Mahwah, NJ: Erlbaum.

Mangelsdorf, S. C., Schoppe, S. J., & Buur, H. (2000). The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 121–140). Mahwah, NJ: Erlbaum.

McHale, J. P., & Irace, K. (2011). Coparenting in diverse family systems. In J. McHale & K. Lindahl (Eds.), *Coparenting: A conceptual and clinical examination of family systems* (pp. 15–37). Washington, DC: American Psychological Association.

McHale, J. P., Kuersten-Hogan, R., & Lauretti, A. (2001). Evaluating coparenting and family-level dynamics during infancy and early childhood: The coparenting and family rating system. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 151–170). Mahwah, NJ: Erlbaum.

Palacios, J., & Brodzinsky, D. (2010). Review: Adoption research: Trends, topics, outcomes. *International Journal of Behavioral Development, 34,* 270–284. doi:10.1177/0165025410362837

Palacios, J., Sanchez-Sandoval, Y., & Leon, E. (2005). Intercountry adoption disruptions in Spain. *Adoption Quarterly, 9,* 35–55. doi:10.1300/J145v09n01_03

Patterson, C. J. (1995). Families of the lesbian baby boom: Parents' division of labor and children's adjustment. *Developmental Psychology, 31,* 115–123. doi:10.1300/J236v04n03_07

Patterson, C. J., & Farr, R. H. (2011). Coparenting among lesbian and gay couples. In J. McHale & K. Lindahl (Eds.), *Coparenting: Theory, research, and clinical applications* (pp. 127–146). Washington, DC: American Psychological Association.

Patterson, C. J., & Hastings, P. D. (2007). Socialization in the context of family diversity. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 433–460). New York: Guilford.

Patterson, C. J., Sutfin, E. L., & Fulcher, M. (2004). Division of labor among lesbian and heterosexual parenting couples: Correlates of specialized versus shared patterns. *Journal of Adult Development, 11,* 179–189. doi:10.1023/B:JADE.0000035626.90331.47

Phillips, B. M., & Lonigan, C. J. (2010). Child and informant influences on behavioral ratings of preschool children. *Psychology in the Schools, 47,* 374–391. doi:10.1002/pits.20476

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Roisman, G. I., Clausell, E., Holland, A., Fortuna, K., & Elieff, C. (2008). Adult romantic relationships as contexts of human development: A multimethod comparison of same-sex couples with opposite-sex dating, engaged, and married dyads. *Developmental Psychology, 44,* 91–101. doi:10.1037/0012-1649.44.1.91

Schoppe, S. J., Mangelsdorf, S. C., & Frosch, C. A. (2001). Coparenting, family process, and family structure: Implications for preschoolers' externalizing behavior problems. *Journal of Family Psychology, 15,* 526–545. doi:10.1037/0893-3200.15.3.526

Schoppe-Sullivan, S. J., Mangelsdorf, S. C., Frosch, C. A., & McHale, J. L. (2004). Associations between coparenting and marital behavior from infancy to the preschool years. *Journal of Family Psychology, 18,* 194–207. doi:10.1037/0893-3200.18.1.194

Solmeyer, A. R., & Feinberg, M. E. (2011). Mother and father adjustment during early parenthood: The roles of infant temperament and coparenting relationship quality. *Infant Behavior & Development, 34,* 504–514. doi:10.1016/j.infbeh.2011.07.006

Stroud, C. B., Durbin, C., Wilson, S., & Mendelsohn, K. A. (2011). Spillover to triadic and dyadic systems in families with young children. *Journal of Family Psychology, 25,* 919–930. doi:10.1037/a0025443

Teubert, D., & Pinquart, M. (2010). The association between coparenting and child adjustment: A meta-analysis. *Parenting: Science and Practice, 10,* 286–307. doi:10.1080/15295192.2010.492040

Van Egeren, L. A. (2004). The development of the coparenting relationship over the transition to parenthood. *Infant Mental Health Journal, 25,* 453–477. doi:10.1002/imhj.20019

Ward, M. (1998). The impact of adoption on the new parents' marriage. *Adoption Quarterly, 2,* 57–78. doi:10.1300/J145v02n02_04

Child Development, July/August 2013, Volume 84, Number 4, Pages 1241–1252

Role Models and the Psychological Characteristics That Buffer Low-Socioeconomic-Status Youth From Cardiovascular Risk

Edith Chen
Northwestern University

William K. Lee, Lisa Cavey,
and Amanda Ho
University of British Columbia

Little is understood about why some youth from low-socioeconomic-status (SES) environments exhibit good health despite adversity. This study tested whether role models and "shift-and-persist" approaches (reframing stressors more benignly while persisting with future optimism) protect low-SES youth from cardiovascular risk. A total of 163 youth, ages 13–16, completed role model interviews and shift-and-persist measures while cholesterol and inflammatory markers, interleukin-6 (IL-6), and C-reactive protein were assessed. Low-SES youth with supportive role models had lower IL-6. Low-SES youth high in shift-and-persist also had lower IL-6. Shift-and-persist partially mediated the interaction of SES and role models on IL-6. Benefits were not found among high-SES youth. Identifying psychological buffers in low-SES youth has implications for health disparities.

Low socioeconomic status (SES) is one of the most robust social determinants of chronic disease in developed countries. For example, individuals living under low-SES circumstances consistently experience greater morbidity and mortality from cardiovascular diseases (CVD), including coronary heart disease and stroke (Kaplan & Keil, 1993; Pollitt, Rose, & Kaufman, 2005). Low-SES is also associated with established risk factors for CVD, including markers of systemic inflammation such as C-reactive protein (CRP) and interleukin-6 (IL-6), and metabolic symptoms such as high cholesterol and obesity (Alley et al., 2006; Dowd, Zajacova, & Aiello, 2010; Karlamangla et al., 2005).

Furthermore, research has identified childhood as an important time during which low-SES seems to have especially potent and lasting effects on health. For example, low childhood SES is associated with greater metabolic risk and insulin resistance during adolescence (Goodman, Daniels, & Dolan, 2007; Goodman, McEwen, Huang, Dolan, & Adler, 2005). In addition, low childhood SES increases the risk of developing stroke and coronary heart disease later in adult life, independent of current SES (Galobardes, Smith, & Lynch, 2006; Miller, Chen, & Parker, 2011).

Resilience in a Low-SES Context

Studies that have linked low-SES to poor health, however, have often left unanswered an important question: Why do some individuals *not* get sick despite facing persistent and severe adversity? Although there are certainly numerous environmental and behavioral factors that explain why low-SES is detrimental to health, including heightened exposure to toxic pollutants, greater engagement in detrimental health behaviors such as smoking, and negative psychological states such as depression (Evans, 2004; Lynch, Kaplan, & Salonen, 1997; Matthews, Gallo, & Taylor, 2010), these factors cannot explain why some individuals thrive despite confronting adverse circumstances such as poverty. This type of thriving has been labeled by researchers as resilience (Masten, 2001; Masten et al., 1988).

Resilience has been extensively discussed within the developmental psychopathology literature (Garmezy, 1985; Masten & Coatsworth, 1998; Rutter, 1987; Werner, 1995). Over the years, this resilience literature has identified key factors at the child (e.g., temperament, emotion regulation), family (e.g., warm, responsive caregiving), and neighborhood (e.g., connections to community adults) levels that buffer children facing adversity from behavioral problems and academic failures (Garmezy,

---

This study was funded by Canadian Institutes of Health Research Grant FRN: 97872. All authors declare no conflicts of interest.

Correspondence concerning this article should be addressed to Edith Chen, Department of Psychology, Northwestern University, Vancouver, BC V6T1Z4, Canada. Electronic mail may be sent to echen@psych.ubc.ca.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0010
DOI: 10.1111/cdev.12037

1985; Luthar, 2006; Masten, 2001; Masten & Coatsworth, 1998; Rutter, 1987). For the most part, however, these studies have focused on psychological adaptation among children (Masten & Coatsworth, 1998; Werner, 1995).

Rarely has the notion of resilience been explored with respect to physical health in children. When considering psychological factors that would promote positive physical health, it is important to identify characteristics that not only are beneficial in the context of adversity but that also have plausible effects on physiological systems that are implicated in disease. For example, traditional models state that low-SES evokes acute physiological stress responses, which over the long term can promote pathogenic processes, such as systemic inflammation, that result in chronic illnesses such as CVD years later (Miller, Chen, & Cole, 2009; Treiber et al., 2003). Hence, psychological factors that buffer the effects of low-SES on health would need to be able to mitigate the physiological responses of these systems to the barrage of stressors that low-SES children face.

Shift-and-Persist

We recently developed a theoretical account of the factors that are protective from a physiological and physical health perspective, identifying a set of psychological characteristics—labeled "shift-and-persist"—that represents a physiologically adaptive approach to dealing with stressors in a low-SES context (Chen & Miller, 2012). We propose that a lifetime of facing constraints with limited options leads those living in a low-SES context to place value on the ability to accept and adjust oneself to stressors that occur in daily life (*shifting*). Shifting entails accommodating the self to stressors through cognitive reappraisals (i.e., reframing the meaning of a stressor in a less threatening manner) and emotion regulation. Coping via efforts to regulate the self may be particularly beneficial under conditions in which stressors are largely uncontrollable and when resources for dealing with stressors are limited (and may in fact be more fruitful than proactively trying to resolve stressors).

At the same time, in this context, successful adaptation entails enduring adversity with strength, finding meaning in difficult situations, and holding on to hope or optimism in the face of adversity (*persisting*). Persisting may be beneficial because it allows low-SES individuals to focus on their futures (rather than on current adversities) and because the search for meaning allows people to maintain hope, particularly when confronting adversity (Updegraff, Silver, & Holman, 2008). We propose that it is the combination of approaches (shifting plus persisting), rather than either one on its own, that will reduce physiological responses to stress acutely among low-SES individuals, and by doing so, over time will mitigate the long-term progression of pathogenic processes that lead to chronic disease (Chen & Miller, 2012).

Developmental Origins of Shift-and-Persist

Developmentally, we propose that one route by which shift-and-persist characteristics emerge is when in the midst of adversity, children find positive role models. Role models refer to important figures who have a significant influence on a child's life (Greenberger, Chen, & Beam, 1998). Role models serve a number of functions related to support and inspiration (Greenberger et al., 1998; Hurd & Zimmerman, 2010; Yancey, Siegel, & McDaniel, 2002).

Role models help to promote shifting-and-persisting in part by facilitating beliefs that others can be trustworthy and dependable, which in turn help shape children's reappraisals of stressful situations. Role models also help orient youth toward their futures, promoting optimism and the discovery of meaning in one's life. In particular, role models can help promote shift-and-persist strategies in youth by modeling the use of shift-and-persist strategies themselves and by coaching youth in the use of these strategies during stressful situations (Chen & Miller, 2012).

Role models (in particular, parents) help children develop emotion regulation strategies beginning early in childhood (Eisenberg, Cumberland, & Spinrad, 1998). However, when it comes to shift-and-persist strategies, we postulate that its development occurs later, typically in adolescence. This is because the abilities required to engage in shift-and-persist (cognitively reframing stressful situations and focusing on one's future in order to maintain hope and optimism) are known to be ones that develop later in childhood (Heckhausen & Schulz, 1995; Nurmi, 1991). For example, 12-year-olds report engaging in significantly more secondary control coping strategies (i.e., adjusting the self) during uncontrollable stressors (medical procedures) compared to 6- and 9-year-olds (Band & Weisz, 1988). Children who had reached the formal operations stage of cognitive development (average age of 14) engaged in significantly more secondary control coping during medical stressors compared to those who were still in the preformal operations stage (average age of 8; Band, 1990). Among children ranging in age from 5 to 11, those who were older were more likely to engage in cognitive distraction strategies and cognitive restructuring (reframing a stressful event in a more positive way) during uncontrollable situations (Altshuler & Ruble, 1989; Curry & Russ, 1985). Taken together, the studies suggest that "shifting" becomes a more widely used coping strategy with increasing age during childhood.

Similarly, with respect to persisting and maintaining hope and optimism by focusing on one's future, older adolescents have been found to be more future oriented than younger adolescents (Nurmi, 1991). They are more likely to articulate goals related to future aspirations such as education, whereas younger adolescents are more likely to describe goals related to leisure activities (Nurmi, Poole, & Kalakoski, 1994). They are also more likely to engage in future planning (Nurmi, 1991). Hence, in this study, we focused on adolescence as an important period for testing the contribution of role models and shift-and-persist to physiological profiles.

Empirical Evidence for Shift-and-Persist

Previous empirical research provides support for the shift-and-persist theory. For example, lifespan theories of control that postulate that when primary control (being able to change the environment so that it fits one's needs and desires) is not possible, secondary control efforts (attempts to bring oneself in line with one's environment) are beneficial (Heckhausen & Schulz, 1995; Heckhausen, Wrosch, & Schulz, 2010). And in fact, active efforts to cope with stressors and being given control over the parameters of a stressor are not beneficial to cardiovascular risk profiles among low-SES individuals (James, Keenan, Strogatz, Browning, & Garrett, 1992; James, Strogatz, Wing, & Ramsey, 1987). Furthermore, finding meaning in life is thought to help individuals cope with traumatic or life-threatening events (Bonanno, Wortman, & Nesse, 2004; Taylor, 1983; Updegraff et al., 2008), and low-SES may represent one circumstance in which broader perspectives on life, such as meaning making and a future orientation are beneficial. Consistent with this notion, low-SES adults who reported greater purpose in life showed lower levels of the inflammatory marker IL-6 (Morozink, Friedman, Coe, & Ryff, 2010).

Two recent empirical studies demonstrated disease-related benefits specific to shift-and-persist. In the first study, the childhood circumstances of a national sample of adults were assessed, and an index of cumulative physiological risk (allostatic load) was obtained. Shift-and-persist scores interacted to predict allostatic load among adults who came from low childhood SES backgrounds. Within this group, the combination of high shift and high persist was associated with the lowest allostatic load scores. In contrast, there was no association between shift-and-persist and allostatic load among those from high childhood SES backgrounds (Chen, Miller, Lachman, Gruenewald, & Seeman, 2012).

Second, in a clinical sample of children diagnosed with asthma, among those low in SES, higher shift-and-persist scores were associated with lower levels of asthma inflammation. In addition, among low-SES children, higher shift-and-persist scores prospectively predicted less asthma impairment (fewer school absences, less rescue inhaler use) 6 months later, controlling for baseline asthma impairment. In contrast, shift-and-persist did not benefit high-SES children with asthma (Chen, Miller, Kobor, & Cole, 2011).

The Present Study

However, no previous studies have yet tested hypotheses about the importance of role models in promoting shift-and-persist, and the implications for physiological risk markers. In addition, indicators of cardiovascular risk have not been examined in the context of shift-and-persist in a younger sample, to document whether the physiological effects of shift-and-persist emerge in adolescence. This study sought to address these two questions in a healthy sample of adolescents from a range of SES backgrounds. We assessed indicators of cardiovascular risk, including two markers of systemic inflammation, CRP and IL-6, high levels of which have been linked to a greater risk of CVD (Ridker, Hennekens, Buring, & Rifai, 2000; Ridker, Rifai, Stampfer, & Hennekens, 2000), as well as total cholesterol, high levels of which also confer risk for CVD (Pekkanen et al., 1990; Ridker, Rifai, Cook, Bradwin, & Buring, 2005). We hypothesized that SES would interact with role model characteristics to predict CRP, IL-6, and cholesterol levels in youth, such that low-SES youth who reported supportive role models would be buffered from elevated inflammatory and metabolic outcomes, whereas role models would not be critical to high-SES youth. We also hypothesized, consistent with previous research, that shift-and-persist strategies would be beneficial to low-SES, but not high-SES, youth in terms of inflammatory and metabolic outcomes. Finally, we hypothesized that the reason why role models would be associated with inflammatory and metabolic outcomes would be because of engagement in shift-and-persist strategies, but only among low-SES youth (a mediated moderation hypothesis).

Method

Participants

A total of 163 families were recruited from Vancouver, British Columbia, Canada from newspaper and transit advertisements. Eligibility criteria included: (a) youth ranging in age from 13 to 16 (all attending secondary school) plus one parent, (b) no chronic illnesses for either parent or youth, (c) free of acute respiratory illness at the time of visit (by parent and youth report), and (d) fluent in English. See Table 1 for descriptive information.

Table 1
Descriptive Sample Information (N = 163)

	Range	%	*M*	*SD*
Age	13–16		14.53	1.03
Gender (% male)		52		
Ethnicity				
Caucasian		47		
Asian		37		
Other		16		
Pubertal stage				
Pre		1		
Early		2		
Mid		25		
Late		59		
Post		13		
Divorced or single (%)		25		
Ever smoked past 6 months (% yes)		9		
Role model (% yes)		83		
Waist:hip ratio	.65–1.15		0.82	0.08
Family savings	0–500,000		19,000	
Shift strategies	3–12		8.06	2.12
Persist strategies	5–24		14.93	3.66
C-reactive protein (mg/L)	.20–8.60		1.12	3.16
IL-6 (pg/ml)	.02–8.32		0.83	1.22
Cholesterol (mg/dl)	8–249	147	34	

Note. Family savings was coded in categories from 1 to 9 (range listed above). On the categorical scale, *M* = 4.90 and *SD* = 2.62; this mean was translated into dollars above to make the number more meaningful in the table.

Procedures

Interested families contacted the lab and were screened for eligibility. Thirty-seven percent of families who called were not eligible, 16% of those who called were not interested in the laboratory visit, and the remainder were scheduled for a lab visit, where consent was obtained from parents and assent for youth. One youth and one caregiver came in for each family. Parents and youth were placed in separate rooms where they completed questionnaires and interviews, described below. A trained phlebotomist drew 50 cc of blood from an antecubital vein in the youth in the same room. Visits typically occurred in the afternoons, and lasted approximately 3 hr (including other interviews that were not relevant to this study), and children and parents each received $75. The protocol was approved by the University of British Columbia Research Ethics Board.

Measures

SES. Socioeconomic resources were measured by asking parents about the amount of assets that their family could easily convert to liquid cash in an emergency (family savings). This measure is recommended by the MacArthur Research Network on Socioeconomic Status and Health (www.macses.ucsf.edu) and is widely used in SES research (Juster, Smith, & Stafford, 1999). In a number of previous studies, we found that resource-based measures of SES such as this one have more robust associations with health-related outcomes in childhood than prestige-based measures (e.g., education), and hence we focused on this construct in this study. Our measure is identical to that used in previous studies (Chen, Cohen, & Miller, 2010).

Shift-and-persist. The tendency to shift oneself in response to stressors was measured in youth using the Positive Thinking scale of the Responses to Stress questionnaire (Connor-Smith, Compas, Wadsworth, Thomsen, & Saltzman, 2000). Three items (e.g., "I told myself that everything would be all right") were queried on a 4-point scale (ranging from *not at all* to *a lot*). These items tap the extent to which individuals try to deal with stressful situations by thinking about them in more positive ways. Items were coded such that higher scores indicated a higher tendency to positively appraise stressful situations (see Table 2). Cronbach's alpha in this sample was .66. This measure has been validated in children and linked to mental health outcomes such as depression (Connor-Smith et al., 2000).

As an indicator of future persistence, a measure of positive thinking about the future was included for youth. The Life Orientation Test taps the extent to which individuals have positive expectations for their future (e.g., "I'm always optimistic about my future"; Scheier, Carver, & Bridges, 1994). This measure consists of six items rated on a 5-point scale (from *strongly disagree* to *strongly agree*). Items were coded such that higher scores indicated higher optimism (see Table 2). Cronbach's alpha in this sample was .73. This measure has been used in children (Chang & Sanna, 2003) and has established links with disease outcomes in adults (Scheier et al., 1999).

To create a total shift-and-persist score, responses to the shift-and-persist measures were first standardized (because they are on different scales), and then summed, as done in our previous study (Chen et al., 2011). Thus, higher scores indicate using a higher combination of both shift-and-persist strategies.

Role models. Youth were interviewed about whether they had a role model, that is, someone in their life whom they looked up to and might want to be like. If youth identified a role model, they were asked what qualities that person possessed that made them a good role model for the participant. Participants responded in an open-ended fashion, with follow-up probes utilized as necessary to clarify the meaning of their responses. Responses regarding the qualities of a good role model were coded by raters who were blind to SES as to whether the role model served a support function, allowing us to focus on one of the critical dimensions of role models identified in previous work (Greenberger et al., 1998; Hurd & Zimmerman, 2010). Specifically, coders were given numerous examples of support statements, and if participants verbalized any type of support function to their role model (e.g., "He helps me with my problems"), then support was coded as yes, with those youth who named other qualities or who did not identify a role model being coded as no. Twenty percent of tapes were coded by two raters for reliability purposes, with a kappa of .84.

Table 2
Shift-and-Persist Items

Shift
1. I told myself that I could get through it, or that I would do better next time.
2. I told myself that everything would be all right.
3. I thought of ways to laugh about it so it wouldn't seem so bad.

Persist
4. In uncertain times, I usually expect the best.
5. If something can go wrong for me, it will.
6. I'm always optimistic about my future.
7. I hardly ever expect things to go my way.
8. I rarely count on good things happening to me.
9. Overall, I expect more good things to happen to me than bad.

Note. Items 1–3 come from the secondary control coping scale of Connor-Smith, Compas, Wadsworth, Thomsen, and Saltzman (2000). Items 4–9 come from Scheier, Carver, and Bridges (1994).

Metabolic and inflammatory measures. Blood was collected through antecubital venipuncture into serum separator tubes. *Total cholesterol* was measured using standard enzymatic techniques using a Hitachi 911 (Kyowa Medex, Japan; interassay coefficient of variance, CV = 0.9%). *IL-6* was measured using a high-sensitivity enzyme-linked immunosorbent assay (ELISA) kit (intraassay CV < 10%; detection threshold = .04 pg/ml; R&D Systems, Minneapolis, MN). *CRP* was measured using a high-sensitivity, chemiluminescent technique on an IMMULITE 2000 (interassay CV = 2.2%; detection threshold = .20 mg/L; Diagnostic Products Corporation, Los Angeles, CA). CRP values were log transformed to normalize their distribution.

Covariates. Covariates were chosen a priori based on variables that are standard for inclusion in psychosocial studies of biological risk markers. These included demographic variables of youth gender and ethnicity, pubertal stage (based on the Peterson Pubertal Development Scale, $\alpha = .77$), a self-report questionnaire that has been validated against the Tanner physical examination evaluation for determining pubertal stage (Petersen, Crockett, Richards, & Boxer, 1988), health behaviors related to smoking (yes or no), and an indicator of obesity, waist–hip ratio. Pubertal stage was used in lieu of age, given the potential impact of pubertal status on metabolic and inflammatory measures that might not be captured solely by including chronological age. Both pubertal stage and waist–hip ratio were standardized within gender, given differences between girls and boys on these measures.

Statistical Analyses

To test associations among SES, role models, shift-and-persist, and metabolic and inflammatory measures, we conducted tests of mediated moderation, as recommended by Muller, Judd, and Yzerbyt (2005). First, we tested whether the effect of role models on metabolic and inflammatory measures was moderated by SES. To do this, multiple regression analyses were conducted in which metabolic or inflammatory scores were predicted from (a) covariates described above, (b) main effect of SES (family savings) and main effect of role model score, and (c) the interaction between SES and role models. Including the interaction term allowed us to test the hypothesis that role models would predict metabolic and inflammatory outcomes only among low-SES youth. Tests of interactions were conducted according to the recommendations of Aiken and West (1991).

Second, we tested whether the potential mediator of this effect (shift-and-persist) also was moderated (i.e., being related in low-SES but not high-SES youth). Analyses were conducted in a parallel fashion to that described above, with metabolic and inflammatory scores predicted from (a) covariates, (b) main effect of SES and main effect of shift-and-persist, and (c) the interaction between SES and shift-and-persist.

Finally, we tested for mediated moderation. Mediated moderation refers to an overall moderation effect (i.e., the effect of role models on metabolic and inflammatory measures being moderated by SES) that is mediated by another variable (i.e., shift-and-persist). That is, we are testing the hypothesis that among low-SES youth, the reason why role models will be associated with better metabolic and inflammatory profiles is because role models facilitate a shift-and-persist approach that has benefits biologically. To test this, we entered the potential mediator into the regression equation simultaneously with SES and the moderator variable (role models). In mediated moderation, the initial moderated effect is reduced when the mediator is included. Thus, our third step was to include the mediator in the first moderator equation described above and to examine any reductions in path coefficients (Muller et al., 2005).

Results

Preliminary Analyses

Descriptive information about the sample is provided in Table 1. SES was not associated with shift-and-persist scores ($p > .2$), the likelihood of having a role model ($p > .2$), the likelihood of having a supportive role model ($p > .6$), age ($p > .6$), pubertal stage ($p > .7$), sex ($p > .3$), ethnicity ($p > .2$), the likelihood of having smoked ($p > .4$), or waist–hip ratio ($p > .2$).

Eighty-three percent of youth identified a role model. The most commonly identified role model was a parent (41%), with the other commonly named types of role models including other family members (20%), celebrities (12%), adult friend (7%), and teacher or coach (6%). Of youth who identified a role model, 25% articulated a support function to their role model. This percentage is somewhat low in part because some participants named role models whom they did not personally know, and in part because our interviews required youth to spontaneously describe their role model as supportive (rather than being asked whether or not their role model was supportive).

In terms of associations of demographic and health behavior variables with other measures, older children had higher levels of CRP ($r = .17$, $p < .05$). Girls were more likely to report a role model and also more likely to report a supportive role model, than boys (χ^2s = 5.24 and 4.54, $ps < .05$). No other associations were significant.

SES, Supportive Role Models, and Metabolic and Inflammatory Measures

After controlling for covariates, there was a significant main effect of SES on IL-6 ($\beta = -.37$, $p < .001$), such that lower SES was associated with greater IL-6 levels. There was no main effect of supportive role models ($\beta = -.13$, $p = .13$). There was, however, a significant interaction of SES with supportive role models ($\beta = .18$, $p < .05$). Figure 1 depicts this interaction graphically. As can be seen in the figure, a supportive role model buffered low-SES youth, but not high-SES youth, from elevated IL-6 levels. No effects were found for cholesterol or CRP (all $ps > .2$).

SES, Shift-and-Persist, and Metabolic and Inflammatory Profiles

We next tested whether shift-and-persist scores were also moderated by SES in their associations with metabolic and inflammatory profiles. After controlling for covariates, there was a significant main effect of SES on IL-6 ($\beta = -.18$, $p < .05$), such that lower SES was associated with greater IL-6 levels. There was no main effect of shift-and-persist ($\beta = -.13$, $p = .14$). There was, however, a significant interaction of SES with shift-and-persist ($\beta = .17$, $p = .05$). Figure 2 depicts this interaction graphically, with lines plotted at ± 1 *SD* of SES. As

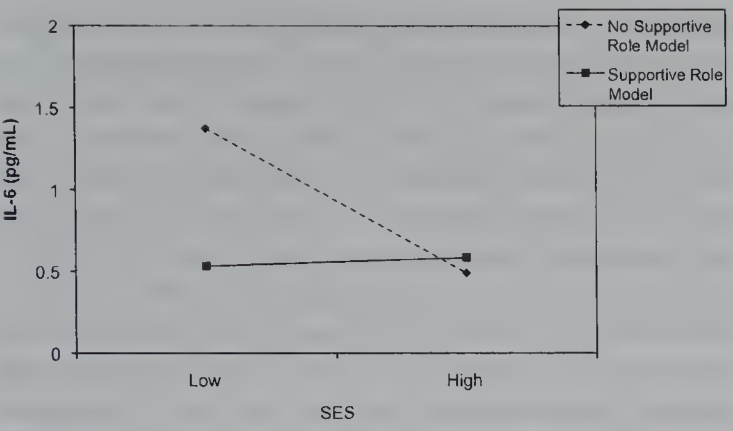


Figure 1. Interaction of socioeconomic status (SES; family savings) by role model characteristics predicting IL-6.

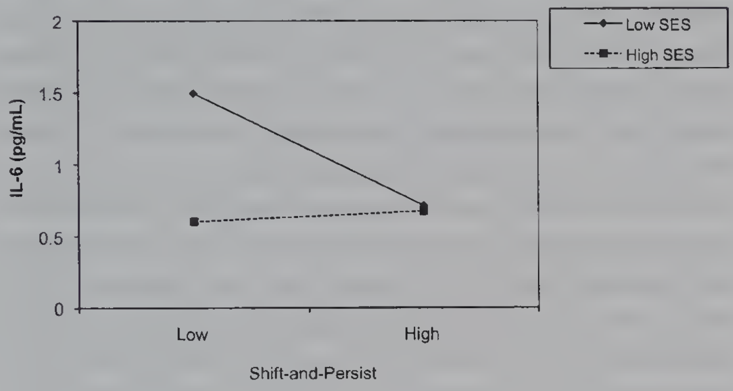


Figure 2. Interaction of socioeconomic status (SES; family savings) by shift-and-persist strategies predicting IL-6. *Note.* Estimated regression lines are plotted at ± 1 *SD* of SES.

can be seen in the figure, the lower the SES, the steeper the association of shift-and-persist with IL-6. That is, among low-SES youth, higher shift-and-persist scores were associated with lower IL-6 levels. Among high-SES youth, there was no relation between shift-and-persist and IL-6. No effects were found for cholesterol or CRP (all *p*s > .4).

Mediated Moderation

Given that we found that SES moderated the relation between supportive role models and IL-6, and that the potential mediator of this relation (shift-and-persist) also was moderated by SES, we tested for mediated moderation by including shift-and-persist in regression equations that tested the effects of SES, supportive role models, and the SES × Role Models interaction on IL-6. After statistically controlling for shift-and-persist, the moderator effect of SES × Role Models became nonsignificant, $\beta = .13$, $p = .17$ (compared to its originally significant value of $\beta = .18$, $p < .05$ with no controls for shift-and-persist). The fact that this moderated relation becomes nonsignificant after controlling for the potential mediator suggests that the tendency to engage in shift-and-persist partially explains why having a supportive role model predicts IL-6 levels among low-SES youth. In addition, we compared the percentage of variance accounted for by the SES × Role Models interaction term with and without shift-and-persist included in the equation. When shift-and-persist was included, the percent variance that the SES × Role Models interaction term accounted for in IL-6 levels was reduced by 48%, suggesting that shift-and-persist explains a substantial proportion of the association between SES × Role Models and IL-6.

Is the Support Function of Role Models Necessary?

We repeated the above analyses utilizing the variable of presence or absence of a role model (regardless of the function the role model serves) to test whether the mere presence of a role model is sufficient to produce these effects. None of the interactions of role model presence or absence with SES was significantly associated with any metabolic or inflammatory measure (*p*s ranging from .26 to .81), suggesting that the support function of role models is an important component for low-SES youth.

Is the Combination of Shift-and-Persist Necessary?

We repeated the above analyses using just shift scores, as well as just persist scores to test whether high levels of either one alone is sufficient to produce a buffering effect among low-SES youth. None of the interactions of shift scores with SES was significantly associated with any metabolic or inflammatory measures (*p*s ranging from .12 to .57). Likewise, none of the interactions of persist scores with SES was significantly associated with any metabolic or inflammatory measures (*p*s ranging from .11 to .51). This suggests that it is the combination of shift-and-persist that is important for predicting metabolic and immune measures in low-SES youth.

Discussion

Our results demonstrate that among low-SES youth, those who had supportive role models exhibited reduced levels of IL-6. Low-SES youth who engaged in shift-and-persist strategies also showed lower levels of IL-6. Shift-and-persist strategies partially mediated the moderated effect of SES by role models on IL-6. In contrast, high-SES youth did not show any physiological benefits of either role models or of engaging in shift-and-persist strategies.

This study is novel in documenting the benefits of role models among low-SES youth in a physical health context, in terms of inflammatory markers. Previous research has revealed beneficial effects of role models primarily within the academic and behavioral contexts. For example, random assignment to formal mentoring programs such as Big Brother Big Sister produced teacher-reported gains in academic performance among a national sample of 9- to 16-year-old youth (Herrera, Grossman, Kauh, & McMaken, 2011). Similarly, other intervention trials that involve mentoring also improve emotional and behavioral, as well as academic outcomes in youth (see DuBois, Holloway, Valentine, & Cooper, 2002, for a review). Observational studies find that natural mentoring relationships (relationships between youth and nonparental adults that are not part of formal programs), or role models, are also associated with fewer behavioral problems, depressive symptoms, and less delinquency, in addition to better grades academically in at-risk youth (Hurd & Zimmerman, 2010; Hurd et al., 2009; Yancey et al., 2002). A few studies have investigated health-related outcomes, primarily in terms of health behaviors, finding that mentors or role models are associated with less substance use and higher levels of physical activity in youth (Beier, Rosenfeld, Spitalny, Zansky, & Bontempo, 2000; DuBois & Silverthorn, 2005; Oman et al., 2004; although associations with substance use have not been found in other studies: Hurd & Zimmerman, 2010; Yancey et al., 2002). However, this study is novel in linking role models to how physiological systems respond to stress among low-SES youth.

The findings in this study linking SES, shift-and-persist, and cardiovascular risk profiles in youth are consistent with previous research on shift-and-persist in children with asthma, as well as in adults (Chen et al., 2011; Chen et al., 2012). These studies all consistently show that there are no main effects of shift-and-persist, only an interaction with SES, indicating that shift-and-persist strategies are not uniformly beneficial, but rather specifically helpful to those who come from low-SES backgrounds. Moreover, in an earlier study, we documented the importance of both shift and persist via a three-way interaction between SES, shift, and persist in predicting cumulative physiological risk in an adult sample (Chen et al., 2012). That is, there was a two-way interaction between shift and persist found in low-SES, but not in high-SES, participants. The two-way interaction was such that the lowest physiological risk was found among those low-SES participants who were high on both shifting and persisting (Chen et al., 2012). Because low-SES individuals on average live under circumstances consisting of more frequent stressors that are more uncontrollable (Brady & Matthews, 2002), an approach that emphasizes shifting oneself (reframing stressors more positively) may be beneficial for slowing down the physiological responses to stressors that over the long term contribute to chronic diseases such as CVD. In addition, maintaining optimism about the future may help provide meaning in life that also mitigates the accumulation of pathogenic processes contributing to chronic disease. Thus, this work suggests the importance of considering context-specific psychological profiles—that is, the notion that certain psychological qualities may not be universally beneficial, but rather are uniquely beneficial, for example, in a low-SES, but not high-SES, context.

Our findings are also more generally consistent with adult studies that have examined psychological characteristics that moderate the effects of SES on health outcomes. For example, low-SES African American adults who are high on John Henryism (the tendency to use active coping efforts for dealing with stressors that are largely uncontrollable) show higher blood pressure and increased risk of hypertension compared to those who are low in SES and low in active coping (James et al., 1987). Among those low in SES, low optimism is associated with higher ambulatory blood pressure and a higher likelihood of hypertension compared to those who were low in SES and high in optimism, or those who were high in SES (Grewen et al., 2000). Similarly, high purpose in life buffered a community sample of adults who were low in SES from high levels of IL-6 (Morozink et al., 2010).

In addition, our findings are consistent with other studies that have examined the benefits of positive relationships with others during childhood among those low in SES. For example, maternal warmth can buffer low-SES individuals from adverse physiological and inflammatory risk profiles in both childhood and adulthood (Chen, Miller, et al., 2011; Evans, Kim, Ting, Tesher, & Shannis, 2007; Miller, Lachman, Chen, Gruenewald, & Seeman, 2011). Also, relevant are studies that have investigated the benefits of social capital (neighborhood connections) in terms of buffering low-SES children from smoking and obesity (Evans & Kutcher, 2011). This study extends this work by specifically focusing on the function of role models for physiological profiles, particularly among low-SES youth.

Because of the cognitive skills required to engage in shift-and-persist, we had hypothesized that these strategies would emerge during adolescence, and hence tested our hypotheses in a sample of 13- to 16-year-olds. We observed in this age group that there were effects of role models among low-SES youth and that the effects of role models were partially mediated by engagement in shift-and-persist strategies. We note that the patterns reported above did not vary by age in our sample (no interactions with age, data not shown); however, the sample did include a fairly restricted age range. Now that this phenomenon has been documented, future research should conduct tests across different development periods to determine at what ages these associations emerge—that is, at what ages it would be most beneficial to have a supportive role model? Also, longitudinal research could test whether role models at earlier ages might facilitate the later development of shift-and-persist. That is, we know that children engage in emotion regulation and coping strategies from a very young age (Eisenberg et al., 1998), that role models can help teach effective strategies to young children (Denham, Mitchell-Copeland, Strandberg, Auerbach, & Blair, 1997), and that adult influences early in life, including maternal sensitivity and harshness, predict later developmental outcomes such as externalizing behavior problems in older children (Bradley & Corwyn, 2007). Thus, could role models who are present at a young age promote shift-and-persist strategies once children become cognitively able to engage in them? Finally, future research should also explore whether there could be effects of negative role models. For example, we know that low-SES parents on average engage in less adaptive parenting behaviors, and that these behaviors in turn can be modeled by children and affect their own well-being outcomes (Conger & Donnellan, 2007). Hence, to what extent are youth susceptible to imitating others who engage in adverse approaches to handling life stressors (thus developing adverse coping strategies from negative role models), and is this susceptibility different for low- versus high-SES adolescents?

We note that this study found evidence of mediated moderation for IL-6, but not for the other physiological markers of CRP or cholesterol. The lack of findings with CRP may have been due in part to CRP levels being quite low during childhood (increasing with age; Ford et al., 2003), and hence there may not have been enough variability in CRP in our age group to find a mediated moderation effect. With respect to cholesterol, it may be the case that markers such as cholesterol are more affected by health behaviors such as diet; in contrast, inflammatory markers such as IL-6 may be more strongly related to psychological pathways such as stress (Segerstrom & Miller, 2004).

Limitations of this study include the cross-sectional design of this study. This is the first study to investigate the contribution of role models to relations among SES, shift-and-persist, and cardiovascular risk, and hence all assessments occurred at one point in time. Thus, we cannot draw any conclusions about causality or directionality. For example, it is possible that those who are already engaging in shift-and-persist strategies are more likely to perceive adults as supportive role models. Future studies that investigate longitudinal associations will allow researchers to better assess the directionality of these relations. Or intervention studies that test the effects of mentoring programs on coping profiles as well as physiological risk profiles would provide stronger conclusions about causality. Second, this study had limited information about role models. The information obtained in this study is in line with questions included in other studies of mentoring or role models (Hurd & Zimmerman, 2010; Hurd et al., 2009); however, given the importance demonstrated in this study of role models, future studies that conduct more in depth assessments of role model characteristics and relationships with youth will be important for ascertaining the critical components, from a health perspective, of youth's relationships with their role models. Third, this study drew on existing measures to derive a shift-and-persist score. While these measures were all psychometrically validated, they were not designed specifically to assess shift-and-persist hypotheses. We are currently in the process of validating a specific shift-and-persist measure, and of testing whether additional components (e.g., future orientation) would better capture this construct. Finally, we point out that our sample in this study may not be that generalizable, given that the study was conducted in Canada and given that the sample was a healthy one with few covariates being associated with inflammatory or metabolic measures.

In sum, among youth who came from low-SES backgrounds, those who had supportive role models showed lower levels of the inflammatory marker IL-6. Low-SES youth who engaged in shift-and-persist strategies also showed lower levels of IL-6, and shift-and-persist partially mediated the relation between role models and IL-6 in low-SES youth. Understanding the specific psychosocial qualities that contribute to physiological resilience among low-SES youth represents one important approach toward interventions to reduce health disparities. The results from this study suggest that it is important to promote shifting-and-persisting in tandem, that is, that neither shifting nor persisting on its own is sufficient to produce physiological benefits, but if the combination can successfully be promoted among low-SES youth, there could be potential benefits for long-term physiological and ultimately, health trajectories in this group. Furthermore, if these are characteristics that some low-SES individuals already possess naturally, they may be ones that would be more plausible to alter through intervention in other low-SES individuals. While there are various approaches that society could take toward reducing health disparities, the ones related to shift-and-persist focus around efforts to help individuals develop self-strategies for dealing with adversity. Alternative approaches might include providing low-SES individuals with basic resources to improve health. Given that no one approach is likely to eliminate the health disparities that are so pervasive in our society, efforts to promote shift-and-persist in conjunction with societal efforts to provide basic economic resources and health services could begin to make a difference in improving the health of children facing adversity early in life.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. London: Sage.

Alley, D. E., Seeman, T. E., Ki Kim, J., Karlamangla, A., Hu, P., & Crimmins, E. M. (2006). Socioeconomic status and C-reactive protein levels in the US population: NHANES IV. *Brain, Behavior, and Immunity*, *20*, 498–504. doi:10.1016/j.bbi.2005.10.003

Altshuler, J. L., & Ruble, D. N. (1989). Developmental changes in children's awareness of strategies for coping with uncontrollable stress. *Child Development*, *60*, 1337–1349.

Band, E. B. (1990). Children's coping with diabetes: Understanding the role of cognitive development. *Journal of Pediatric Psychology*, *15*, 27–41.

Band, E. B., & Weisz, J. R. (1988). How to feel better when it feels bad: Children's perspectives on coping with everyday stress. *Developmental Psychology*, *24*, 247–253.

Beier, S. R., Rosenfeld, W. D., Spitalny, K. C., Zansky, S. M., & Bontempo, A. N. (2000). The potential role of an adult mentor in influencing high-risk behaviors in adolescents. *Archives of Pediatrics and Adolescent Medicine*, *154*, 327–331.

Bonanno, G. A., Wortman, C. B., & Nesse, R. M. (2004). Prospective patterns of resilience and maladjustment during widowhood. *Psychology and Aging*, *19*, 260–271. doi:10.1037/0882-7974.19.2.260

Bradley, R. H., & Corwyn, R. F. (2007). Externalizing problems in fifth grade: Relations with productive activity, maternal sensitivity, and harsh parenting from infancy through middle childhood. *Developmental Psychology*, *43*, 1390–1401.

Brady, S. S., & Matthews, K. A. (2002). The effect of socioeconomic status and ethnicity on adolescents' exposure to stressful life events. *Journal of Pediatric Psychology*, *27*, 575–583.

Chang, E. C., & Sanna, L. J. (2003). Experience of life hassles and psychological adjustment among adolescents: Does it make a difference if one is optimistic or pessimistic? *Personality and Individual Differences*, *34*, 867–879.

Chen, E., Cohen, S., & Miller, G. E. (2010). How low socioeconomic status affects 2-year hormonal trajectories in children. *Psychological Science*, *21*, 31–37. doi:10.1177/0956797609355566

Chen, E., & Miller, G. E. (2012). "Shift-and-persist" strategies: Why being low in socioeconomic status isn't always bad for health. *Perspectives on Psychological Science*, *7*, 135–158.

Chen, E., Miller, G. E., Kobor, M. S., & Cole, S. W. (2011). Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. *Molecular Psychiatry*, *16*, 729–737. doi:10.1038/mp. 2010.53

Chen, E., Miller, G. E., Lachman, M. E., Gruenewald, T. L., & Seeman, T. E. (2012). Protective factors for adults from low childhood socioeconomic circumstances: The benefits of shift-and-persist for allostatic load. *Psychosomatic Medicine*, *74*, 178–186.

Chen, E., Strunk, R. C., Trethewey, A., Schreier, H. M., Maharaj, N., & Miller, G. E. (2011). Resilience in low-socioeconomic-status children with asthma: Adaptations to stress. *Journal of Allergy and Clinical Immunology*, *128*, 970–976. doi:10.1016/j.jaci.2011.06.040

Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology*, *58*, 175–199. doi:10.1146/annurev.psych.58.110405.085551

Connor-Smith, J. K., Compas, B. E., Wadsworth, M. E., Thomsen, A. H., & Saltzman, H. (2000). Responses to stress in adolescence: Measurement of coping and involuntary stress responses. *Journal of Consulting and Clinical Psychology*, *68*, 976–992.

Curry, S. L., & Russ, S. W. (1985). Identifying coping strategies in children. *Journal of Clinical Child Psychology*, *14*, 61–69.

Denham, S. A., Mitchell-Copeland, J., Strandberg, K., Auerbach, S., & Blair, K. (1997). Parentalc ontributions to preschoolers' emotional competence: Direct and indirect effects. *Motivation and Emotion*, *21*, 65–86.

Dowd, J. B., Zajacova, A., & Aiello, A. E. (2010). Predictors of inflammation in U.S. children aged 3–16 years.

American Journal of Preventive Medicine, *39*, 314–320. doi:10.1016/j.amepre.2010.05.014

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, *30*, 157–197.

DuBois, D. L., & Silverthorn, N. (2005). Natural mentoring relationships and adolescent health: Evidence from a national study. *American Journal of Public Health*, *95*, 518–524. doi:10.2105/AJPH.2003.031476

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental socialization of emotion. *Psychological Inquiry*, *9*, 241–273.

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, *59*, 77–92.

Evans, G. W., Kim, P., Ting, A. H., Tesher, H. B., & Shannis, D. (2007). Cumulative risk, maternal responsiveness, and allostatic load among young adolescents. *Developmental Psychology*, *43*, 341–351.

Evans, G. W., & Kutcher, R. (2011). Loosening the link between childhood poverty and adolescent smoking and obesity: The protective effects of social capital. *Psychological Science*, *22*, 3–7.

Ford, E. S., Giles, W. H., Myers, G. L., Rifai, N., Ridker, P. M., & Mannino, D. M. (2003). C-reactive protein concentration distribution among US children and young adults: Findings from the National Health and Nutrition Examination Survey, 1999–2000. *Clinical Chemistry*, *49*, 1353–1357.

Galobardes, B., Smith, G. D., & Lynch, J. W. (2006). Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Annals of Epidemiology*, *16*, 91–104.

Garmezy, N. (1985). Stress-resistant children: The search for protective factors. In J. E. Stevenson (Ed.), *Recent research in developmental psychopathology* (pp. 213–233). Oxford, England: Pergamon.

Goodman, E., Daniels, S. R., & Dolan, L. M. (2007). Socioeconomic disparities in insulin resistance: Results from the Princeton School District Study. *Psychosomatic Medicine*, *69*, 61–67.

Goodman, E., McEwen, B. S., Huang, B., Dolan, L. M., & Adler, N. E. (2005). Social inequalities in biomarkers of cardiovascular risk in adolescence. *Psychosomatic Medicine*, *67*, 9–15.

Greenberger, E., Chen, C., & Beam, M. R. (1998). The role of "very important" nonparental adults in adolescent development. *Journal of Youth and Adolescence*, *27*, 321–343.

Grewen, K., Girdler, S. S., West, S. G., Bragdon, E., Costello, N., & Light, K. C. (2000). Stable pessimistic attributions interact with socioeconomic status to influence blood pressure and vulnerability to hypertension. *Journal of Women's Health & Gender-Based Medicine*, *9*, 905–915. doi:10.1089/152460900750020946

Heckhausen, J., & Schulz, R. (1995). A life-span theory of control. *Psychological Review*, *102*, 284–304.

Heckhausen, J., Wrosch, C., & Schulz, R. (2010). A motivational theory of life-span development. *Psychological Review*, *117*, 32–60. doi:10.1037/a0017668

Herrera, C., Grossman, J. B., Kauh, T. J., & McMaken, J. (2011). Mentoring in schools: An impact study of Big Brothers Big Sisters school-based mentoring. *Child Development*, *82*, 346–361. doi:10.1111/j.1467-8624.2010.01559.x

Hurd, N., & Zimmerman, M. (2010). Natural mentors, mental health, and risk behaviors: A longitudinal analysis of African American adolescents transitioning into adulthood. *American Journal of Community Psychology*, *46*, 36–48. doi:10.1007/s10464-010-9325-x

Hurd, N. M., Zimmerman, M. A., & Xue, Y. (2009). Negative adult influences and the protective effects of role models: A study with urban adolescents. *Journal of Youth and Adolescence*, *38*, 777–789. doi:10.1007/s10964-008-9296-5

James, S. A., Keenan, N. L., Strogatz, D. S., Browning, S. R., & Garrett, J. M. (1992). Socioeconomic status, John Henryism, and blood pressure in black adults. The Pitt County Study. *American Journal of Epidemiology*, *135*, 59–67.

James, S. A., Strogatz, D. S., Wing, S. B., & Ramsey, D. L. (1987). Socioeconomic status, John Henryism, and hypertension in Blacks and Whites. *American Journal of Epidemiology*, *126*, 664–673.

Juster, F. T., Smith, J. P., & Stafford, F. (1999). The measurement and structure of household wealth. *Labour Economics*, *6*, 253–275.

Kaplan, G. A., & Keil, J. E. (1993). Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation*, *88*, 1973–1998.

Karlamangla, A. S., Singer, B. H., Williams, D. R., Schwartz, J. E., Matthews, K. A., Kiefe, C. I., et al. (2005). Impact of socioeconomic status on longitudinal accumulation of cardiovascular risk in young adults: The CARDIA Study (USA). *Social Science & Medicine*, *60*, 999–1015.

Luthar, S. S. (2006). Resilience in development: A synthesis of research across five decades. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation* (2nd ed., pp. 739–795). New York: Wiley.

Lynch, J. W., Kaplan, G. A., & Salonen, J. T. (1997). Why do poor people behave poorly? Variation in adult health behaviors and psychosocial characteristics by stages of the socioeconomic lifecourse. *Social Science and Medicine*, *44*, 809–819.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, *56*, 227–238.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments. *American Psychologist*, *53*, 205–220.

Masten, A. S., Garmezy, N., Tellegen, A., Pellegrini, D. S., Larkin, K., & Larsen, A. (1988). Competence and stress in school children: The moderating effects of individual and family qualities. *Journal of Child Psychiatry*, *29*, 745–764.

Matthews, K. A., Gallo, L. C., & Taylor, S. E. (2010). Are psychosocial factors mediators of socioeconomic status and health connections? A progress report and blueprint for the future. *Annals of the New York Academy of Sciences, 1186*, 146–173. doi:10.1111/j.1749-6632.2009.05332.x

Miller, G., Chen, E., & Cole, S. W. (2009). Health psychology: Developing biologically plausible models linking the social world and physical health. *Annual Review of Psychology, 60*, 501–524.

Miller, G. E., Chen, E., & Parker, K. J. (2011). Psychological stress in childhood and susceptibility to the chronic diseases of aging: Moving toward a model of behavioral and biological mechanisms. *Psychological Bulletin, 137*, 959–997. doi:10.1037/a0024768

Miller, G. E., Lachman, M. E., Chen, E., Gruenewald, T. L., & Seeman, T. E. (2011). Pathways to resilience: Maternal nurturance as a buffer against childhood poverty's effects on metabolic syndrome at midlife. *Psychological Science, 22*, 1591–1599.

Morozink, J. A., Friedman, E. M., Coe, C. L., & Ryff, C. D. (2010). Socioeconomic and psychosocial predictors of interleukin-6 in the MIDUS national sample. *Health Psychology, 29*, 626–635. doi:10.1037/a0021360

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863.

Nurmi, J. E. (1991). How do adolescents see their future? A review of the development of future orientation and planning. *Developmental Review, 11*, 1–59.

Nurmi, J. E., Poole, M. E., & Kalakoski, V. (1994). Age differences in adolescent-future-oriented goals, concerns, and related temporal extension in different sociocultural contexts. *Journal of Youth and Adolescence, 23*, 471–487.

Oman, R. F., Vesely, S., Aspy, C. B., McLeroy, K. R., Rodine, S., & Marshall, L. (2004). The potential protective effect of youth assets on adolescent alcohol and drug use. *American Journal of Public Health, 94*, 1425–1430.

Pekkanen, J., Linn, S., Heiss, G., Suchindran, C. M., Leon, A., Rifkind, B. M., et al. (1990). Ten-year mortality from cardiovascular disease in relation to cholesterol level among men with and without preexisting cardiovascular disease. *New England Journal of Medicine, 322*, 1700–1707. doi:10.1056/NEJM199006143222403

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence, 17*, 117–133.

Pollitt, R. A., Rose, K. M., & Kaufman, J. S. (2005). Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: A systematic review. *BMC Public Health, 5*, 7. doi:10.1186/1471-2458-5-7

Ridker, P. M., Hennekens, C. H., Buring, J. E., & Rifai, N. (2000). C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *New England Journal of Medicine, 342*, 836–843.

Ridker, P. M., Rifai, N., Cook, N. R., Bradwin, G., & Buring, J. E. (2005). Non-HDL cholesterol, apolipoproteins A-I and B100, standard lipid measures, lipid ratios, and CRP as risk factors for cardiovascular disease in women. *Journal of the American Medical Association, 294*, 326–333. doi:10.1001/jama.294.3.326

Ridker, P. M., Rifai, N., Stampfer, M. J., & Hennekens, C. H. (2000). Plasma concentration of interleukin-6 and the risk of future myocardial infarction among apparently healthy men. *Circulation, 101*, 1767–1772.

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry, 57*, 316–331.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*, 1063–1078.

Scheier, M. F., Matthews, K. A., Owens, J. F., Schulz, R., Bridges, M. W., Magovern, G. J., et al. (1999). Optimism and rehospitalization following coronary artery bypass graft surgery. *Archives of Internal Medicine, 159*, 829–835.

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin, 130*, 601–630.

Taylor, S. E. (1983). Adjustment to threatening events: A theory of cognitive adaptation. *American Psychologist, 38*, 1161–1173.

Treiber, F. A., Kamarck, T., Schneiderman, N., Sheffield, D., Kapuku, G., & Taylor, T. (2003). Cardiovascular reactivity and development of preclinical and clinical disease states. *Psychosomatic Medicine, 65*, 46–62.

Updegraff, J. A., Silver, R. C., & Holman, E. A. (2008). Searching for and finding meaning in collective trauma: Results from a national longitudinal study of the 9/11 terrorist attacks. *Journal of Personality and Social Psychology, 95*, 709–722. doi:10.1037/0022-3514.95.3.709

Werner, E. E. (1995). Resilience in development. *Current Directions in Psychological Science, 4*, 81–85.

Yancey, A. K., Siegel, J. M., & McDaniel, K. L. (2002). Role models, ethnic identity, and health-risk behaviors in urban adolescents. *Archives of Pediatrics & Adolescent Medicine, 156*, 55–61.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1253–1268

Informants' Traits Weigh Heavily in Young Children's Trust in Testimony and in Their Epistemic Inferences

Jonathan D. Lane, Henry M. Wellman, and Susan A. Gelman
University of Michigan

This study examined how informants' traits affect how children seek information, trust testimony, and make inferences about informants' knowledge. Eighty-one 3- to 6-year-olds and 26 adults completed tasks where they requested and endorsed information provided by one of two informants with conflicting traits (e.g., honesty vs. dishonesty). Participants also completed tasks where they simultaneously considered informants' traits and visual access to information when inferring their knowledge and trusting their testimony. Children and adults preferred to ask and endorse information provided by people who are nice, smart, and honest. Moreover, these traits influenced the knowledge that young children attributed to informants. Children younger than 5 years of age reported that people with positive traits were knowledgeable even when they lacked access to relevant information.

Much of the knowledge that children and adults hold is derived from information others provide. From others, we acquire information ranging from the mundane—today's date or the weather—to information that cannot be easily observed or assessed—the presence of germs, the existence of God, or the nature of the afterlife (Harris & Koenig, 2006). Given young children's limited experience with the world, testimony is a particularly valuable source of information for them (Gelman, 2009; Harris, 2007). But not all people are sources of reliable information. Some people unknowingly offer inaccurate information because of their own misperceptions or ignorance. Others intentionally (perhaps malevolently) provide false information. Because of this variability in informants' trustworthiness, it is important for children to be selective in whom they choose to learn from. The current study examines how informants' traits—specifically, their smartness, kindness, and honesty—affect how children seek information, trust testimony, and make inferences about informants' knowledge.

Preschool-age children appreciate that people possess traits—intangible, stable dispositions that guide thinking and behavior (for a review, see Heyman, 2009)—and indeed young children can use others' traits to predict their psychological experiences and behavior (Heyman & Gelman, 1999, 2000; Liu, Gelman, & Wellman, 2007). How children consider informants' traits when acquiring knowledge is a critical issue to address because the people from whom children typically learn are people with whom children have extensive experience—parents, friends, teachers—and thus children have substantial information about those individuals' stable characteristics. Harris (2007) argues that children create "cognitive profiles" of informants:

> They form a global impression of each individual, regarding some as more epistemically trustworthy than others. We assume that this global impression regarding any given informant is based on some kind of aggregated metric. Information about the informant's past inaccuracy, ignorance, uncertainty, or apparent idiosyncrasy is fed into that profile. (Harris, 2007, p. 138)

Funding for this research was provided by National Institute of Child Health and Human Development Grant HD022149 to Henry Wellman and by Rackham School of Graduate Studies Research Grant to Jonathan Lane. We are very grateful to the children and parents who participated in this study and to the administration, teachers, and staff at the University of Michigan Children's Centers and Go Like the Wind Montessori School. We appreciate Becca Leider for her help interviewing children, and Daniel Blumer, Jennifer Elledge, and Amanda Nagrotsky for their help entering and checking data. The manuscript benefited from feedback provided by Vikram Jaswal and other members of International Max Planck Research School on the Life Course, and benefited from comments and suggestions provided by four anonymous reviewers. A portion of this research was presented at the annual meeting of the Association for Psychological Science, May 2010.

Correspondence concerning this article should be addressed to Jonathan D. Lane, Harvard Graduate School of Education, Cambridge, MA 02138. Electronic mail may be sent to jonathan_lane@gse.harvard.edu.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0011
DOI: 10.1111/cdev.12029

Much of the recent work on children's trust in testimony examines how one potentially trait-related characteristic—informants' accuracy—affects the way that children learn from others. In one standard task, children are shown two characters, typically hand puppets or real people on video (Clément, Koenig, & Harris, 2004; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005), who differ in their demonstrated accuracy (e.g., their correct or incorrect naming of common objects) and children choose to *ask* one character to learn something new (e.g., the name of a novel object). Then both characters offer conflicting information (testimony) and children choose which character's testimony they believe by *endorsing* one or the other. When learning the names of novel objects, 3- and 4-year-olds prefer to ask previously accurate informants (Koenig et al., 2004). Four-year-olds additionally prefer to endorse object labels provided by accurate informants (Clément et al., 2004; Koenig et al., 2004).

Although 3-year-olds often do not preferentially endorse accurate informants' testimony in these tasks, those who keep better track of informants' prior accuracy more often endorse testimony provided by more accurate informants (Koenig et al., 2004). Three-year-olds also selectively endorse informants when presented more data about informants' tendencies. In cases where young children see one informant is consistently accurate across three or four instances, and the other informant is inaccurate across three or four instances, even 3-year-olds trust testimony provided by the previously accurate informant (Birch, Vauthier, & Bloom, 2008; Pasquini, Corriveau, Koenig, & Harris, 2007; Scofield & Behrend, 2008). However, if both informants are wrong at least once (e.g., one is correct 25% of the time whereas the other is correct 75% of the time), 3-year-olds no longer demonstrate selective trust; 4-year-olds do (Pasquini et al., 2007). Thus, 3-year-olds in particular seem to require consistent evidence about informants' tendencies (in this case, their accuracy) in order to develop a profile of an "accurate" or "inaccurate" informant.

Being accurate or inaccurate is not a trait per se, and this is something that even preschoolers realize. For example, preschoolers are willing to trust recently inaccurate informants if their inaccuracies stem from ignorance or false beliefs (Nurmsoo & Robinson, 2009; Robinson & Nurmsoo, 2009)—in these cases, it is clear that inaccuracy does not indicate a broad, trait-like mental shortcoming. However, systematic accuracy might be related to underlying traits, such as intelligence. Given their inductive potential (Heyman, 2009) traits may exert a particularly powerful influence on children's trust in certain informants; however, this possibility is largely unaddressed in the literature. The aforementioned studies collectively show that the stronger the evidence about informants' systematic accuracy, the more heavily children weigh that tendency when trusting testimony. Thus, characters' enduring, underlying traits may be especially effective in eliciting selective trust from young children. These considerations framed our methods in the current study. Children saw two informants who differed in the accuracy of their prior testimony (as in Koenig et al., 2004; Koenig & Harris, 2005). Importantly, children also received additional information about these informants' traits that might influence the quality of their testimony; for example, the informants were assigned the trait labels "smart" and "not smart," we described what both of those labels mean, then reminded children that the informants possess those traits and checked that children understood.

In addition to smartness, we examined whether and when young children take into account other traits, specifically kindness and honesty. Kindness (being mean or nice) was included because it is a trait understood even by young preschoolers; they appreciate that nice and mean people engage in different behaviors and have different motives (Heyman & Gelman, 1999, 2000). Although, for adults, kindness does not necessarily reflect the quality of one's epistemic states, because it has an emotional component and because it may influence children's global evaluations of individuals, kindness may nonetheless influence children's selective trust. Honesty has strong emotional appeal as well, and is also a characteristic that young children are beginning to understand. Preschoolers are sensitive to statements that are lies and in some studies even 3-year-olds understand lying as a psychological process (Siegal & Peterson, 1996, 1998). In addition to its potential emotional appeal, honesty may directly reflect the quality of information someone is likely to produce, so it is a particularly important trait to consider when assessing testimony. Conceivably, whether informants are generally dishonest (or honest) could influence young children's willingness to trust their testimony.

In the prior work examining the influence of informants' kindness and honesty on children's trust, Mascaro and Sperber (2009, Studies 2 and 3) had a single informant (a frog puppet) provide young children with erroneous testimony about the location of candy. In one study, the informant was described as a "Big liar! It always tells lies." In the other study, the informant was described as "Very mean! It does not want you to find the sweet!" In both studies, children were presented two boxes and the informant made a claim that the candy was in one of the boxes. It was not until age 4 years in Study 2 (and still later in Study 3) that children avoided the box indicated by the deceptive frog. However, these results may not reflect children's consideration of informants' honesty in particular. In their Study 2, children were warned about the informant ("Watch out!") as well as told he was a liar, so it is not clear if 4-year-olds' appropriate choices reflect the warning, the lying, or both. Moreover, these results likely reflect young children's more general difficulty disregarding claims or instructions given by a *single* informant, regardless of his characteristics. This tendency is particularly strong when informants provide spoken information (Jaswal, Croft, Setia, & Cole, 2010; Reed, Pien, & Rothbart, 1984; Strommen, 1973; Vanderbilt, Liu, & Heyman, 2011). Prior studies seem to indicate that children's sensitivity to characteristics of informants (e.g., accuracy, age) is most clearly demonstrated when children choose between the conflicting testimony of *two* informants (Jaswal & Neely, 2006; Koenig et al., 2004; Pasquini et al., 2007). Indeed, when Mascaro and Sperber (2009) described a single deceptive informant as "mean" in their Study 3, preschoolers still followed his testimony, even though 3-year-olds preferred the testimony of a nice informant (*rather* than the testimony of a mean informant) in their Study 1. Thus, almost no research has addressed children's appreciation for informants' honesty when learning from others, and the available data may underestimate children's understanding. Here, we employ a standard testimony paradigm—having children choose between honest and dishonest informants who offer conflicting information—to examine whether young children weigh informants' honesty when learning from others.

In summary, we comprehensively compare the influence of several different traits—smartness, kindness, and honesty—on how children seek and endorse information provided by others. All of these traits are worth considering when learning from others, but the developmental point at which children begin to take them into consideration may differ.

Of note, young children's performance in prior studies varies depending on the nature of the task: including whether there is a single or multiple informants (Mascaro & Sperber, 2009) and, of particular relevance to the current study, if children are to *ask* versus *endorse* informants. For example, as Koenig et al. (2004) demonstrated, 3-year-olds prefer to ask a recently accurate (vs. inaccurate) person for information, but are equally likely to endorse the two informants' subsequent testimony. This may be because the two tasks differ in what they assess, with the former potentially measuring a global attraction (or aversion) to an informant or a desire to affiliate with one informant over the other, and the latter a more direct measure of children's consideration of the actual *information* that people offer. This distinction is especially important to consider when informants have positive and negative traits. Thus, we include both types of questions in the current study: *ask* (Who will you ask for information, X or Y?) and *endorsement* (Do you believe X's or Y's testimony?).

Preference for Information Versus Desire to Affiliate

The affiliation possibility, just discussed, deserves further attention. Some have challenged the notion that children's choice of informants and information on standard testimony tasks actually reflects children's consideration of informants' epistemic states (e.g., Lucas & Lewis, 2010). One alternative possibility is that children's responses to both ask and endorsement tasks reveal nothing more than a general attraction to certain people, rather than an appreciation for the quality of information the people have to offer. Several studies in which accurate informants are pitted against inaccurate informants demonstrate that children's preference for testimony is not based on a general attraction to accurate informants (e.g., Birch et al., 2008; Jaswal & Neely, 2006). However, informants in the current study have strongly valenced traits (e.g., meanness and niceness) which may more strongly influence informants' global attractiveness to children. Thus, in addition to comparing children's performance across the ask and endorsement tasks, we also directly assess children's general attraction to each informant by having children report whether they would like to be friends with each informant. This allowed us to assess relations between children's general attraction to certain people and their tendency to ask and endorse information provided by those same people.

Even if not driven solely by a general attraction to certain informants, perhaps initially children weigh informants' more global characteristics (e.g., traits) over more proximate factors (e.g., informants' current epistemic states) when deciding whose information to trust. Alternatively, as we discuss next, informants' more enduring, global qualities may influence children's inferences about informants' epistemic states, and this in turn may influence whose testimony children trust.

Considering Informants' Knowledge

One goal of research on children's trust in testimony is to understand when children appreciate that certain people possess better *knowledge* than others. Theoretically, the nature of informants' traits may not matter if they are ignorant—if they *lack* the knowledge that a learner wants to acquire. An ignorant person's testimony may be untrustworthy, regardless of whether he is intelligent, kind, or honest. However, traits carry strong inductive potential and may lead children astray when making inferences about epistemic states. Because children often form global evaluations of individuals (as good or bad) across multiple domains (Alvarez, Ruble, & Bolger, 2001; Cain, Heyman, & Walker, 1997; Stipek & Daniels, 1990), positive traits may suggest to young children that informants are generally knowledgeable even when they are not. So, in examining their understanding of the relevance of various traits for testimony, it is important also to assess children's understanding of informants' knowledge.

Prior work indicates that young children are appropriately sensitive to informants' knowledge in certain scenarios. For example, children as young as 3 years of age are less willing to believe the testimony of someone who is ignorant because he lacks visual access to relevant information (Robinson, Champion, & Mitchell, 1999). If faced with a scenario in which one informant accurately labels common objects and the other informant consistently says "I do not know," 3-year-olds prefer to ask the accurate informant for the label of a new object, and 4-year-olds prefer to ask *and endorse* labels provided by the (thus far) accurate informant (Koenig & Harris, 2005). Four-year-olds are also more accepting of information provided by informants who are confident in their knowledge (Moore, Bryant, & Furrow, 1989). For example, they believe information provided by someone who says, "I know it is in the red box" versus someone who says, "I guess it is in the red box" (see also Jaswal & Malone, 2007; Sabbagh & Baldwin, 2001). Building on these beginnings, more information is needed about how informants' access to knowledge affects children's trust in testimony, particularly in relation to their appreciation for traits such as smartness, which may also convey to young children information about knowledge.

In the current study, our approach to these issues is to assess how children simultaneously consider informants' traits *and* their access to context-specific information. In real life, these factors do not always converge: A "good" person is sometimes ignorant; a "bad" person may be knowledgeable (e.g., an accessory to a crime). This interplay between traits and knowledge may be problematic for preschoolers, and understanding how children negotiate such situations may provide important information about the development of trust in testimony. For example, if preschoolers prefer the testimony of well-informed people (e.g., Robinson et al., 1999), they may continue to favor their testimony even if those informants possess negative traits. In contrast, certain traits may overinfluence children's inferences about the quality of informants' knowledge and also their subsequent decisions about who to trust. Moreover, how children simultaneously weigh informants' traits and knowledge access may vary depending upon the trait. For example, smartness may be weighted more heavily than niceness, because smartness links more directly to knowledge states. To examine these as-yet-untested possibilities, we asked children to decide between the testimony of positive-trait informants who are poorly informed versus negative-trait informants who are well informed.

In summary, the current study is designed to address four questions: (a) How do informants' traits—specifically, their honesty, kindness, and smartness—influence young children's choice of informants and trust in their testimony? (b) Is children's preference for informants with certain traits based solely on a desire to affiliate with those informants? (c) How do informants' traits influence children's inferences about their knowledge? and (d) How do children simultaneously weigh informants' traits and their knowledge when trusting testimony?

Method

Participants

Participants included 94 children between 3 and 6 years of age. Of these children, six were not included in our analyses because they ended the interview session early (three 3-year-olds, two 4-year-olds, and one 5-year-old) and another seven (six 3-year-olds and one 4-year-old) were excluded because they failed to remember which trait corresponded to each of the characters. The final sample consisted of 81 children, including twenty-five

3-year-olds (14 girls, 11 boys; range = 34.57–48.77 months, M = 41.91), thirty-two 4-year-olds (12 girls, 20 boys; range = 49.20–59.83 months, M = 54.58), and twenty-four 5- and 6-year-olds (11 girls, 13 boys; range = 61.03–82.50 months, M = 69.77). Children were interviewed for approximately 20 min in a quiet, familiar room in their school. Children attended schools in a Midwestern university town, were primarily White, and were of middle to upper-middle socioeconomic status. When parents gave permission, interviews were audio recorded. To assess what a mature pattern of responses would be for our tasks, 26 adults (15 women, 11 men; age range = 18 to 79 years, M = 34.77) completed an online version of the study with identical characters, scenarios, and questions.

Measures and Procedure

Children heard three pairs of stories (presented in story books) in which two pictured characters exhibited contrasting traits: Honest–dishonest, nice–mean, and smart–not smart. The six stories and pictures are presented in the Appendix. To maximize children's understanding of these characters' traits as enduring, stable qualities and not fleeting characteristics, children were given multiple converging pieces of information about each character. For each character, children were given a trait label (e.g., "Sam is smart"), description of the trait (e.g., "Sam is a boy who gives right answers"), scenario (e.g., Sam correctly identifying that an animal is a frog), and reminder about the trait (e.g., "Remember, Sam is smart"). Interviewers were careful to use a similar intonation during the presentation of all characters. The order in which the trait pairs were presented, the order in which characters were presented within pairs, and characters' appearance/name within pairs were counterbalanced. For each trait pair, children were shown laminated drawings of both characters and were asked to recall which trait corresponds to each character (e.g., "Which one is smart, Sam or Noah?" and "Which one is not smart, Sam or Noah?"). These memory checks were administered twice for each pair: immediately after a pair was introduced and half-way through the set of questions for that pair. Corrective feedback was provided when necessary. Children who were incorrect for both memory checks for any of the traits were excluded from analyses.

Following the presentation of each character, children were asked whether they wanted to be friends with him (affiliation questions). For each of the three trait pairs, children were given three tasks (numbered below as 1, 2, 3), each containing two elements. (1) Object naming: (a) asking one of the two informants about the name of a novel object, (b) endorsing one informant's testimony about the object's name; (2) Box contents: (a) asking one of the two informants about the contents of a box that both characters looked inside, (b) endorsing one informant's testimony about the box's contents; (3) Knowledge: (a) judging which character is knowledgeable about the contents of a closed box when only *one* character (the negative-trait character) has looked in the box, (b) endorsing testimony from one of the two informants about the contents of the closed box. The order of the three tasks was counterbalanced, and within trait pairs, location/presentation order (left/first or right/second) was counterbalanced. For all of these tasks, characters were presented as drawings on laminated cards (the same cards used for the memory checks described above). Characters were voiced by the experimenter and their actions (e.g., approaching and opening boxes) were illustrated by the experimenter manually moving the drawings. Additional details of these procedures are provided below.

Affiliation. To assess children's desire to affiliate with each character, following each character's introduction, children were asked, "Would you like to be friends with (character's name)?"

Object naming. Children were shown a picture of a novel object situated between the two informants and were asked, "Do you know what this is called?" The majority of the children said "No." Children were then told, "You can ask (first informant) or (second informant) what this is called. Who do you want to ask?" Both characters offered a label and children were asked: "What do you think this is called, (first informant's label) or (second informant's label)?"

Box contents (open box). Children were shown an open box (with the opening facing away from the child) and were asked, "Do you know what is inside this box?" Most children replied "No." Children were then told, "(First informant) and (second informant) are going to look in the box." Both characters leaned over the box and children were told, "Let us find out what is inside. Who do you want to ask, (first informant) or (second informant)?" Both characters offered an answer (the name of a familiar object) and children were asked, "What do you think is inside, (first informant's answer) or (second informant's answer)?"

Knowledge (closed box). For each trait pair, children also received a task in which the negative-trait character (dishonest, mean, not smart) looked inside a box, and the positive-trait character (honest, nice, smart) did not. So the valence of the trait conflicted with the informants' access to knowledge. For each trait-pair, children were shown a closed box and were asked, "Do you know what is inside this closed box?" and "Can you see inside this box?" Most children answered "No" to both questions. Next, children were told: "(Positive-trait character) and (negative-trait character) have never ever seen inside this closed box before. Now, (positive-trait character) picks up the closed box, walks, and puts it down over here. (Negative-trait character) opens the box, looks inside, and then closes the box." Children were then asked the focal question: "Who knows what is inside the box, (positive-trait character) or (negative-trait character)?" For each trait pair, the order in which the characters acted on the box was counterbalanced.

In order to examine how children simultaneously weigh informants' traits and their access to knowledge, we had children choose between testimony offered by the two informants: negative-trait informants with access to knowledge and positive-trait informants without access to knowledge. Thus, following children's decisions about the informants' knowledge in this task, both characters made claims about the box's supposed contents (offering the name of a familiar object), and children were asked, "What do you think is inside, (first informant's answer) or (second informant's answer)?"

Results

Preliminary analyses revealed no gender differences in children's performance on our tasks. Thus, data from girls and boys were combined in the following analyses.

Asking Informants for Information

We first report children's decisions to seek information from certain people. Children's decisions to ask certain informants for information were comparable when learning the name of the novel object and when learning the contents of the open box that both informants looked inside. Thus, children's performance was averaged across the two tasks. For each trait pair, children could earn a score ranging from 0% (never asked the positive-trait character) to 100% (asked the positive-trait character on both tasks). Figure 1 depicts children's decisions to seek information from each of the positive-trait informants. An initial 3 (trait) × 3 (age group) analysis of variance (ANOVA) examining children's decisions to ask the informants with positive traits (i.e., the honest, nice, and smart informants) revealed a significant effect of age group, $F(2, 78) = 5.17$, $p < .01$, but no significant effect of trait or interaction between age group and trait. Five- and six-year-olds ($M = 95\%$, $SD = 13\%$) asked the positive-trait informants significantly more often than did 3-year-olds ($M = 77\%$, $SD = 24\%$, $p < .01$ according to Tukey's honestly significant difference [HSD] pairwise comparisons), and marginally more often than did 4-year-olds ($M = 83\%$, $SD = 23\%$, $p = .07$ according to Tukey HSD pairwise comparisons).

Analyses within each age group revealed that when presented informants who possessed stable underlying traits, children as young as 3 years preferred to ask the nice (vs. mean) informant: 3-year-olds, $t(24) = 3.36$, $p < .01$; 4-year-olds, $t(31) = 7.00$, $p < .001$; and 5- to 6-year-olds, $t(23) = 12.69$, $p < .001$. More novel to our study was the inclusion of a pair of informants who differed in their smartness and another pair who differed in their honesty. We found that 3-year-olds, $t(24) = 3.38$, $p < .01$; 4-year-olds, $t(31) = 6.17$, $p < .001$; and 5- to 6-year-olds, $t(23) = 15.91$, $p < .001$, asked the smart (vs. not smart) informant above chance, and that children as young as 3 years also preferred to seek information from the honest (vs. dishonest) informant: 3-year-olds, $t(24) = 5.20$, $p < .001$; 4-year-olds, $t(31) = 4.72$, $p < .001$; and 5- to 6-year-olds, $t(23) = 15.91$, $p < .001$. Thus, informants' smartness, kindness, and honesty all significantly influence who young children choose to learn from, and the influence of these traits increases throughout the preschool years.

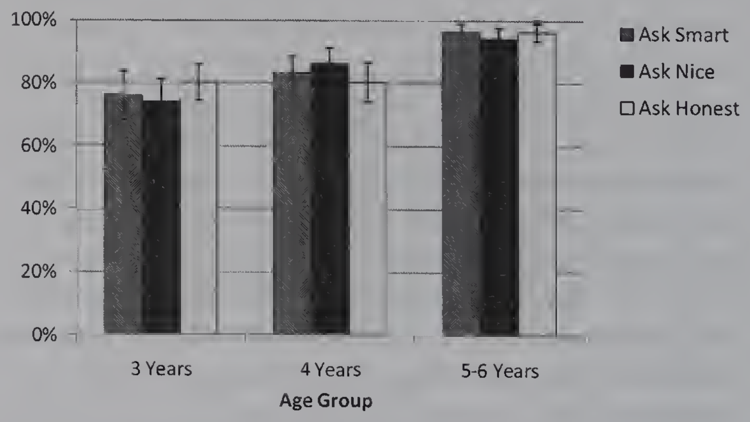


Figure 1. Percentage of decisions to ask the positive-trait characters (smart, nice, and honest) for information, per age group.

Endorsing Informants' Testimony

As noted in the Introduction, asking one informant over another may be a measure of a desire to affiliate as much as an evaluation of testimony and particularly if children are given trait information about the informants. Thus, it is important to also consider endorsement—a task that might more directly gauge children's appreciation for the quality of the information rather than for the informants themselves. For each trait pair, for both object-naming and box-contents tasks, after children chose to ask one informant, both informants offered testimony (labels for the novel object or names of the object contained in the box) and children endorsed one informant's testimony. Children's decisions about which informant's testimony they trusted were comparable in both tasks, so we again averaged children's performance across the two tasks. For each trait pair, children could earn a score ranging from 0% (never endorsed positive-trait character's testimony) to 100% (endorsed positive-trait character's testimony on both tasks).

Figure 2 illustrates children's decisions to endorse testimony provided by each of the positive-trait informants. An initial 3 (trait) × 3 (age group) ANOVA assessed children's endorsement of testimony provided by the positive-trait informants. This analysis revealed a significant effect of age group, $F(2, 78) = 10.91$, $p < .001$, but no significant effect of trait or interaction between age group and trait. Five- and 6-year-olds ($M = 96\%$, $SD = 14\%$) endorsed the positive-trait informants' testimony significantly more often than did 3-year-olds ($M = 67\%$, $SD = 27\%$, $p < .001$ according to Tukey HSD pairwise comparisons) and 4-year-olds ($M = 74\%$, $SD = 24\%$, $p < .01$ according to Tukey HSD pairwise comparisons).

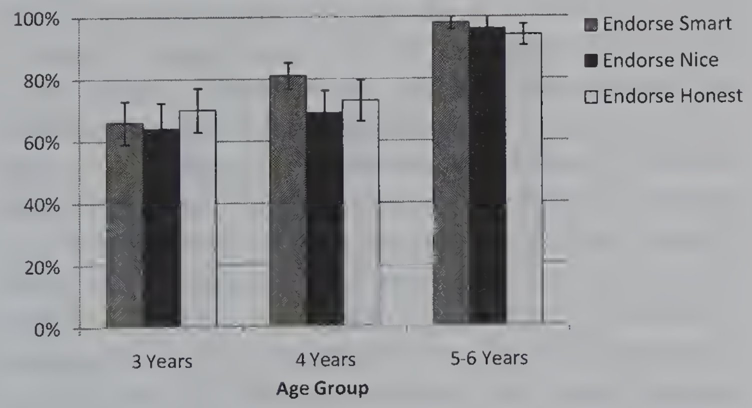


Figure 2. Percentage of decisions to endorse information provided by the positive-trait characters (smart, nice, and honest), per age group.

Considering first nice versus mean informants, 4-year-olds, $t(31) = 2.55$, $p < .05$, and 5- to 6-year-olds, $t(23) = 11.00$, $p < .001$, endorsed the information provided by the nice informant above chance, and 3-year-olds tended in that direction, $t(24) = 1.66$, $p = .11$. More novel to our study was the inclusion of informants who differed in their smartness and honesty. When smartness was presented as a stable underlying trait, children as young as 3 years of age endorsed testimony provided by the smart (vs. not smart) informant above chance: 3-year-olds, $t(24) = 2.32$, $p < .05$; 4-year-olds, $t(31) = 7.19$, $p < .001$; and 5- to 6-year-olds, $t(23) = 23.00$, $p < .001$. Moreover, children as young as 3 years of age preferred testimony provided by the honest (vs. dishonest) informant above chance: (3-year-olds, $t(24) = 2.83$, $p < .01$; 4-year-olds, $t(31) = 3.48$, $p < .01$; and 5- to 6-year-olds, $t(23) = 12.69$, $p < .001$. Though 3-year-olds did not endorse the nice informant's testimony above chance, their preference for his testimony did not differ significantly from their preference for the smart or honest informant's testimony, $F(2, 23) = .17$, *ns*. In summary, informants' smartness, honesty, and (to a lesser extent) their kindness influence children's trust in testimony. These effects emerge as early as age 3 and become more pronounced throughout the preschool years.

If children's decisions to ask an informant were influenced by their general attraction to him without regard for his epistemic states, this might be especially the case for a trait such as niceness (kindness). Indeed, with regard to the least epistemically relevant trait, kindness, children were more likely to ask the nice informant (85%) than they were to endorse his subsequent testimony (75%), $t(80) = 2.15$, $p < .05$. In contrast, for smartness and honesty, children's decisions about whose testimony to endorse were equivalent in frequency to their decisions about who to ask—children were equally likely to ask the smart informant (84%) and to endorse his testimony (81%), $t(80) = .74$, *ns*, and were equally likely to ask the honest informant (85%) and to endorse his testimony (78%), $t(80) = 1.43$, *ns*. These data are relevant for an "affiliation" hypothesis, which we examined more directly by asking children whether they would like to be friends with each of the characters.

Affiliation with Informants

As shown in Figure 3, 4- through 6-year-olds wanted to be friends with all of the positive-trait characters (smart, nice, honest), but not with any

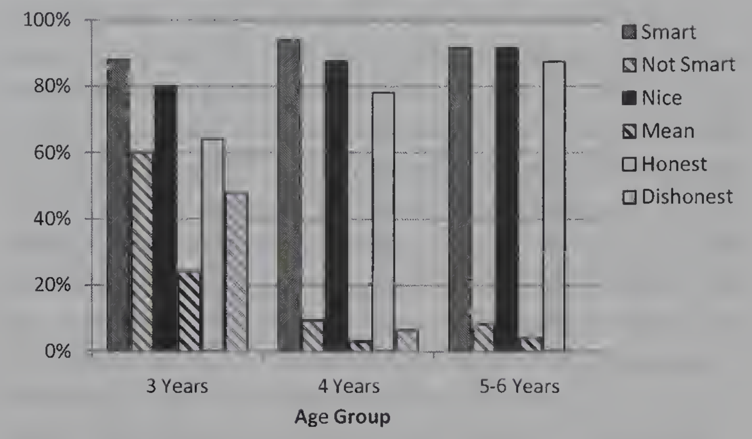


Figure 3. Percentage of children who chose to be friends with each of the characters, per age group.

of the negative-trait characters (not smart, mean, dishonest). Three-year-olds exhibited a different pattern. Like the older children, 3-year-olds wanted to be friends with the nice character but not with the mean character. However, in contrast to older children, roughly half of the 3-year-olds wanted to be friends with informants who were dishonest or not smart. Further, there was no statistical difference in 3-year-olds' decisions to be friends with the honest versus the dishonest character, $\chi^2(1) = 1.50$, *ns*.

This pattern of results provides a direct test of the affiliation hypothesis. Suppose children's trust merely reflects a desire to affiliate with or a general attraction to certain informants. Then, given 3-year-olds' overwhelming desire to affiliate with the nice (and not the mean) informant and the minimal difference in their desire to affiliate with the smart (vs. not smart) informant or with the honest (vs. dishonest) informant, they should selectively trust nice over mean informants but show a different pattern of trust for the smart–not smart pair and the honest-dishonest pair. However, the 3-year-olds were equally willing to ask and to endorse information from the nice, smart, and honest characters: ask, $F(2, 23) = .38$, *ns*, and endorse: $F(2, 23) = .17$, *ns*.

More generally, across all participants, a sizable number of children ($n = 19$) reported *not* wanting to be friends with the honest character (often indicating that they chose to not be friends with him because he violated a social norm by breaking a window). If children's trust simply reflects a desire to affiliate with or general attraction to the honest (vs. dishonest) character, then these children should trust the honest character less than children who *did* want to be his friend ($n = 62$). However, these two groups of children were equally likely to ask the honest informant, $t(79) = .96$, *ns*, and to endorse his testimony, $t(79) = 1.50$, *ns*. Moreover, children's preference to ask and endorse information provided by the honest informant was above chance even for children who did not want to be friends with him: ask, $t(18) = 3.64$, $p < .01$, and endorse, $t(18) = 2.11$, $p < .05$.

Finally, for each age group, we conducted correlations between children's desire to be friends with positive-trait informants (a score of 0–3) and negative-trait informants (a score of 0–3) on the one hand, and their tendency to ask and endorse the positive-trait informants on the other. Among 3-year-olds, many of whom chose to be friends with positive- *and* negative-trait informants, friendship choices were unrelated to their decisions to ask and to endorse the positive-trait informants. Among older children, who typically only wanted to be friends with positive-trait informants, friendship choices were also unrelated to their endorsement of the positive-trait informants' testimony. In these analyses, friendship choice was only related to older children's decisions about who to *ask* for information: Children who wanted to be friends with fewer negative-trait characters were more likely to *ask* the positive-trait informants for information: 4-year-olds, $r(30) = -.56$, $p < .001$, and 5- to 6-year-olds, $r(22) = -.66$, $p < .001$. In summary, it is unlikely that children's trust in testimony in the current study solely reflects an overwhelming general attraction to these positive characters or a general aversion to the negative characters.

Considering Informants' Knowledge

Potentially, others' traits may influence more than just children's decisions about whose testimony they should seek and endorse. Given their strong inductive potential, traits may influence the very knowledge children attribute to informants. We examined this possibility with the *knowledge* task in which the negative-trait characters looked inside a closed box (so they had knowledge of the box's contents) and the positive-trait characters did not look in the box (so they were ignorant about the box's contents). Thus, the quality of knowledge conflicted with the valence of the trait. Children were asked which character knows what is inside the box. Across the entire sample, correct attributions of knowledge to the negative-trait characters did not vary between the three trait pairs (Cochran's $Q = .267$, *ns*). Thus, we averaged across all three trait pairs, so children could earn scores ranging from 0% (never attributed knowledge to negative-trait characters) to 100% (attributed knowledge to all three negative-trait characters). Figure 4 depicts children's correct attributions of

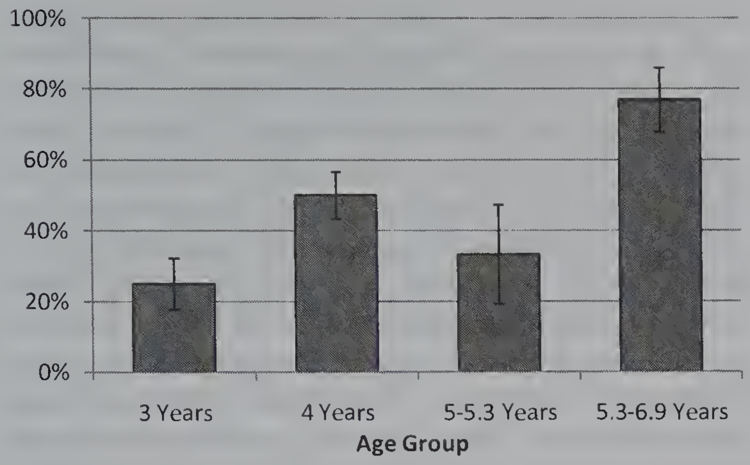


Figure 4. Attributions of knowledge to the negative-trait characters, who had perceptual access to relevant information.

knowledge to the negative-trait characters. A one-way ANOVA revealed a significant difference in attributions of knowledge between the three age groups, $F(2, 78) = 5.91$, $p < .01$. Four-year-olds and 5- to 6-year-olds attributed knowledge to the negative-trait characters significantly more than did 3-year-olds ($ps = .05$ and .003, respectively, according to Tukey HSD pairwise comparisons). Across the whole sample, there was an age-graded increase in correct attributions of knowledge to the negative-trait characters, $r(79) = .34$, $p < .01$.

In prior studies, when characters do not possess traits, children as young as 3 years attribute knowledge to characters who perceive information, but not to those who do not perceive the information (e.g., Pratt & Bryant, 1990). However, our 3-year-olds often reported, above chance, that the positive-trait characters knew the contents of the closed box, even though those characters never looked inside the box, $t(24) = -3.39$, $p < .01$. Four-year-olds performed at chance, $t(31) = 0.00$, *ns*. The oldest group as a whole performed accurately across 62% of the trials, but this was no different from chance, $t(23) = 1.46$, *ns*. Note that our oldest group included children ranging in age from 5 years to nearly 7 years. Inspection of the data suggested that the very oldest children were consistently accurate on this task. To confirm this, we examined the oldest 20% of our sample ($n = 16$; age range = 63.50–82.50 months). These children accurately attributed knowledge to the negative-trait characters for 77% of the trials, and their performance was significantly above chance, $t(15) = 3.01$, $p < .01$.

This initial portion of the knowledge task concerned simply children's attribution of knowledge to the characters. A further important question concerns children's endorsement of the *testimony* provided by the positive-trait/ignorant versus negative-trait/knowledgeable characters. Following children's decisions about the informants' knowledge, for each trait pair, both informants provided testimony about the closed box's contents and children chose to endorse information provided by one of the informants. One child, a 3-year-old, did not provide an answer to this question for the honest–dishonest pair. An initial one-way ANOVA assessing the frequency with which children endorsed the positive-trait (though ignorant) informants' testimony revealed no significant difference between the three age groups, $F(2, 78) = 1.75$, *ns* (3-year-olds: 57%; 4-year-olds: 71%; 5- to 6-year-olds: 71%).

Children's trust in the positive-trait informants' testimony depended upon whether they attributed knowledge to those informants. This pattern was especially clear for nice–mean and honest–dishonest pairs of informants. Children who correctly identified that the mean informant was knowledgeable were less likely to endorse the nice informant's subsequent testimony, $\chi^2(1) = 5.19$, $p < .05$) and children who identified that the dishonest informant was knowledgeable were less likely to endorse the honest informant's testimony, $\chi^2(1) = 6.47$, $p = .01$. Analyzed separately, children who incorrectly attributed knowledge to the nice and honest informants endorsed their subsequent testimony at levels above chance (nice: 77%, binomial $p < .001$; honest: 77%, binomial $p < .001$). In contrast, children who were correct in *not* granting knowledge to the nice and honest informants did not endorse their testimony above chance (nice: 53%, binomial $p = .871$; honest: 50%, binomial $p = 1.00$). A different pattern emerged for the smart-not-smart pair. Children were equally trusting of the smart character's testimony regardless of whether they understood that he lacked the critical knowledge, $\chi^2(1) = 1.20$, *ns*; collectively children preferred his testimony 69% of the time, significantly above chance (binomial $p < .001$).

In summary, trust in informants with positive traits was mitigated among children who could identify when those informants lacked relevant knowledge. However one trait in particular—informants' smartness—continued to weigh heavily in children's trust in testimony, even among children who could identify when a smart informant lacked critical knowledge.

Adult Judgments

In order to validate our intuitions about what a mature pattern of responses would be for these tasks, 26 adults answered identical questions. In the object-naming and the box-contents tasks, these adults overwhelmingly preferred to ask the positive-trait informants (nice: *M* = 100%, *SD* = 0%; smart: *M* = 98%, *SD* = 10%; honest: *M* = 100%, *SD* = 0%) and overwhelmingly endorsed their testimony (nice: *M* = 100%, *SD* = 0%; smart: *M* = 96%, *SD* = 14%; honest: *M* = 100%, *SD* = 0%). For the knowledge task, adults accurately attributed knowledge to the negative-trait informants on 85% of the trials, significantly above chance, $t(25) = 5.35$, $p < .001$. When faced with trusting the testimony of either the positive-trait/ignorant or negative-trait/knowledgeable informants, adults trusted the positive-trait/ignorant informants at chance (*M* = 49%, *SD* = 37%).

These data from adults also provided a point of comparison for our oldest children, to determine if our 5- and 6-year-olds were exhibiting an adult-like pattern of responses. Indeed, for the standard ask and endorsement questions (when learning about the novel object and about the contents of the open box), adults' responses were not significantly different from the 5- and 6-year-olds' responses. Thus their data looked essentially identical to that from the 5- to 6-year-olds in Figures 1 and 2. Moreover, for the knowledge task, adults' attribution of knowledge to the negative-trait informants was not significantly different from the 16 oldest children's knowledge attributions. The only difference between the oldest children and adults was that adults were less trusting of the positive-trait/ignorant informants (*M* = 49%, *SD* = 37%) than were the oldest children (*M* = 71%, *SD* = 36%), $t(48) = 2.13$, $p < .05$.

Discussion

Much of our knowledge is acquired from others—parents, friends, experts, and teachers. These individuals not only differ in their knowledge but also in many other regards, such as their appearance, age, and personality. The extent to which children weigh some of these factors when learning from others has recently received enormous empirical attention (for reviews, see Gelman, 2009; Harris, 2007). In the current study, we assessed the influence of informants' traits on how children seek information, trust testimony, and make inferences about informants' knowledge. We additionally examined how both informants' traits and their access to knowledge simultaneously influence children's trust in testimony.

Our interest in the influence of informants' traits on children's epistemic trust was motivated, in part, by recent studies which collectively demonstrate that, if presented enough consistent data about informants' tendencies, children as young as 3 years selectively trust informants based upon their recent accuracy (e.g., Birch et al., 2008; Pasquini et al., 2007; Scofield & Behrend, 2008). Some have reasoned that children use informants' recent accuracy to construct "cognitive profiles" regarding informants' epistemic trustworthiness (Harris, 2007). If provided enough converging information about an informant's accuracy, conceivably children may even begin to infer a trait (perhaps smartness), which may then influence children's subsequent epistemic trust. In the current study, we took a direct approach to assessing the influence of informants' traits on children's trust—we used a paradigm in which characters were assigned trait labels, the labeled traits were described, characters behaved in accordance with their trait labels, and children were reminded about the characters' traits. We found that children as young as 3 years of age, tend to ask and endorse information provided by smart informants rather than informants who are not smart. We also found that preschoolers prefer to ask and endorse information provided by informants who are nice—a less epistemically relevant trait. Although 3-year-olds were as trusting of nice informants as they were the other positive-trait informants, children were not significantly more trusting of the nice (vs. mean) informant until 4 years, and overall children preferred to ask as opposed to endorse the nice informant.

More novel, we assessed children's sensitivity to another trait that is of particular import when considering the quality of someone's testimony—honesty. The influence of informants' honesty on children's knowledge acquisition has received surprisingly little attention in the literature, with the few available studies suggesting that children do not begin to consider informants' honesty until age 4 or later (Mascaro & Sperber, 2009; Vanderbilt et al., 2011). However, using the same experimental paradigm employed for the other traits—which included two informants who provided conflicting testimony—we found that children as young as 3 years of age also prefer to ask and endorse testimony provided by honest (as opposed to dishonest) informants. What likely accounts for discrepancies between findings is that prior studies required that children disregard information or instructions provided by a single informant, tasks that children may have difficulty with regardless of informants' specific traits (Jaswal et al., 2010; Reed et al., 1984; Strommen, 1973). Thus, prior studies appear to have underestimated children's consideration of informants' honesty. In general, informants' traits—especially smartness and honesty—can weigh heavily when young children consider the quality of testimony.

Thus, the current results demonstrate that others' traits influence children's epistemic trust. We demonstrated this by pitting characters with traits of contrasting valence—for example, mean versus nice—against one another. With our procedures, it is unclear whether children trusted positive-trait informants, *distrusted* negative-trait informants, or both. Future studies could more meticulously assess the driving forces behind the effects found for each of the trait pairs. For example, positive- and negative-trait informants can be pitted against neutral informants, to determine whether children's responses are driven entirely by the positive-trait informants, entirely by the negative-trait informants, or by a combination of the two.

One possible interpretation of our results is that children chose informants and testimony based upon a general characterization of the smart, honest, and nice informants as equally positive, without considering their particular traits. However, the variability in 3-year-olds' willingness to affiliate with each of the six informants suggests that children did not consider the three positive-trait informants one and the same. The affiliation data also allowed us to examine relations between preschoolers' testimony judgments and their more general attraction or aversion to informants. Several studies pitting accurate informants against inaccurate informants have demonstrated that children's preference for testimony cannot be attributed to children's general attraction to accurate informants (e.g., Birch et al., 2008; Jaswal & Neely, 2006). The results of the current study likewise demonstrate that a general attraction to informants with certain traits is not solely responsible for children's trust in their testimony. Over half of the 3-year-olds wanted to be friends with the informant who was not smart, yet they typically chose to ask and endorse information provided by the smart informant. Moreover, 3-year-olds were equally willing to be friends with the honest and dishonest informants, yet they preferred to ask the honest informant for information and they favored his testimony; and children were equally willing to trust information provided by the honest informant regardless of whether they wanted to be his friend. For all age groups, decisions to affiliate with informants were uncorrelated with children's endorsement of informants' testimony, also indicating that a general desire to affiliate was not solely responsible for children's trust in testimony. However, the affiliation questions and testimony questions had different formats—for each pair, children could choose to be friends with one, both, or neither of the informants, but had to trust the testimony of just one informant. To further examine relations between children's affiliation preferences and trust in testimony, future studies can use a forced-choice format where children choose to be friends with just one of the two informants.

An important focus of our research was children's consideration not only of informants' traits but of their knowledge as well. Here, our most intriguing finding was that traits can influence the very knowledge that young children attribute to informants. Indeed, 3-year-olds overwhelmingly used the valence of informants' traits, rather than informants' perceptual access to relevant information, to infer informants' knowledge. This was true for a trait that is often associated with the quality of one's knowledge (smartness) as well as two traits that seem more morally relevant and less epistemically relevant (i.e., not typically associated the quality of individuals' knowledge), honesty and kindness. Only the oldest children (5- and 6-year-olds) and adults consistently differentiated between the valence of informants' traits and the quality of their knowledge. Thus, although 3- and 4-year-olds are fully aware of the link between perceiving information and knowing that information (e.g., Pratt & Bryant, 1990), they sometimes disregard people's perceptual access to information when those people have certain traits, and instead (inaccurately) use the valence of those traits to guide epistemic inferences. Others' traits carry great inductive potential when considering their intentions, emotions, and behavior (Heyman, 2009), and young children appear to use trait information even when reasoning about others' knowledge.

In a study with seemingly contrasting results, Brosseau-Liard and Birch (2010) found that 4-year-olds did not attribute greater general knowledge (e.g., knowledge of words) or situation-specific knowledge (e.g., "Who knows where I put my books?") to a recently accurate character. Likewise, Brosseau-Liard and Birch (2011) demonstrated that preschoolers show no preference for the testimony of accurate (vs. inaccurate) informants when both informants report on the contents of a box that neither informant has looked in. The results of these studies are likely a product of children only having been given information about informants' recent accuracy (which may or may not reflect an underlying trait), rather than the rich and more direct information that we provided about our informants' traits. Young children in particular have difficulty predicting people's future behavior based upon their prior trait-related behavior, but can do so when people are assigned trait labels (Heyman & Gelman, 1999, 2000; Liu et al., 2007). We anticipated that this may also be the case with regard to children's use of trait information to guide their epistemic trust and epistemic inferences. Thus, in addition to engaging in trait-consistent behavior, our informants were assigned trait labels and we described to children what those labels meant.

Turning to our results on how children simultaneously consider informants' knowledge access and traits when trusting testimony, we found that children who reported (incorrectly) that nice and honest informants were knowledgeable about key information (when in fact those informants had not even perceived the information) tended to endorse those informants' testimony. In contrast, children who understood that those informants were *not* knowledgeable were less trusting of their testimony. Intriguingly, a different pattern emerged for the smart informant—children trusted his testimony even when they understood that he was not knowledgeable. Thus, when learning from others, young children may weigh informants' general intelligence more strongly than informants' specific knowledge, especially in cases where a smart informant is pitted against an informant who is not smart.

The early tendency to attribute knowledge to individuals with positive qualities should be considered when interpreting other findings on young children's preference for certain informants' testimony. For example, Corriveau and Harris (2009) demonstrated that 3- and 4-year-olds prefer to ask and endorse information provided by teachers they know well, and continue to trust these teachers even if they had just provided inaccurate information. It is not until age five that children prefer information from unfamiliar yet accurate teachers. Corriveau and Harris provide four potential interpretations of their findings, and settle upon a "halo effect" account—children judge familiar people as positive on multiple dimensions, including likeability, competence, and trustworthiness. Such a halo effect may indeed be at work (see Brosseau-Liard & Birch, 2010), and we would add that the quality of informants' knowledge (including situation-specific knowledge) may be included within such a halo.

Our data also lend support to a complementary account of Corriveau and Harris's (2009) results: given extended experience with their teachers, children appreciate that they possess certain traits, specifically intelligence—teachers are typically knowledgeable, provide accurate information, can figure out new problems, and so on. As we argued earlier, young children in particular require much detail about someone's tendencies in order to infer their traits, and thus a few recent instances of the familiar teacher's inaccuracy may have not been enough information for the youngest children to understand that she possessed some fundamental cognitive limitation, especially when they have so much prior information about the familiar teacher's competence.

In summary, informants' traits can factor strongly into how young children seek and trust testimony. Preschoolers prefer to ask and endorse information provided by people who are nice, smart, and honest, and this is especially the case for the latter two traits. This preference increases throughout the preschool years and is not attributable to a more general desire to affiliate with positive-trait individuals. Moreover, these traits influence the very knowledge that preschoolers attribute to informants. In certain circumstances, young children consider people with positive traits to be knowledgeable even if they clearly do not have access to relevant information. This misattribution of knowledge decreases throughout the preschool years, but is still apparent in 5-year-olds. In young children's evaluation of testimony, informants' epistemically relevant traits can weigh particularly heavily. Indeed, as one example, children continue to trust the testimony of smart (over not smart) informants even when they realize that such informants lack key knowledge.

References

Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development*, *72*, 1409–1425. doi:10.1111/1467-8624.00356

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*, 1018–1034. doi:10.1016/j.cognition.2007.12.008

Brosseau-Liard, P. E., & Birch, S. A. J. (2010). "I bet you know more and are nicer too!": What children infer from others' accuracy. *Developmental Science*, *13*, 772–778. doi:10.1111/j.1467-7687.2009.00932.x

Brosseau-Liard, P. E., & Birch, S. A. J. (2011). Epistemic states and traits: Preschoolers appreciate the differential informativeness of situation-specific and person-specific cues to knowledge. *Child Development*, *82*, 1788–1796. doi:10.1111/j.1467-8624.2011.01662.x

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domains. *Social Development*, *6*, 53–75. doi:10.1111/j.1467-9507.1997.tb00094.x

Clément, F., Koenig, M. A., & Harris, P. L. (2004). The ontogenesis of trust. *Mind and Language*, *19*, 360–379. doi:10.1111/j.0268-1064.2004.00263.x

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*, 426–437. doi:10.1111/j.1467-7687.2008.00792.x

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, *60*, 115–140. doi:10.1146/annurev.psych.59.103006.093659

Harris, P. L. (2007). Trust. *Developmental Science*, *10*, 135–138. doi:10.1111/j.1467-7687.2007.00575.x

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*, 505–524. doi:10.1111/j.1467-8624.2006.00886.x

Heyman, G. D. (2009). Children's reasoning about traits. In P. J. Bauer (Ed.), *Advances in child development and behavior* (Vol. 37, pp. 105–143). New York: Academic Press.

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inferences. *Child Development*, *70*, 604–619. doi:10.1111/1467-8624.00044

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology*, *77*, 1–19. doi:10.1006/jecp. 1999.2555

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*, 1541–1547. doi:10.1177/0956797610383438

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development*, *8*, 263–283. doi:10.1080/15248370701446392

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*, 757–758. doi:10.1111/j.1467-9280.2006.01778.x

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 694–698*, doi:10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277. doi:10.1111/j.1467-8624.2005.00849.x

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development*, *78*, 1543–1558. doi:10.1111/j.1467-8624.2007.01082.x

Lucas, A. J., & Lewis, C. (2010). Should we trust experiments on trust? *Human Development*, *53*, 167–172. doi:10.1159/000320044

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*, 367–380. doi:10.1016/j.cognition.2009.05.012

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, *60*, 167–171.

Nurmsoo, E., & Robinson, E. J. (2009). Children's trust in previously inaccurate informants who were well or poorly informed: When past errors can be excused. *Child Development*, *80*, 23–27. doi:10.1111/j.1467-8624.2008.01243

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226. doi:10.1037/0012-1649.43.5.1216

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*, 973–982. doi:10.1111/j.1467-8624.1990.tb02835.x

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly*, *30*, 131–147.

Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, *35*, 535–546. doi:10.1037/0012-1649.35.2.535

Robinson, E. J., & Nurmsoo, E. (2009). When do children learn from unreliable speakers? *Cognitive Development*, *24*, 16–22. doi:10.1016/j.cogdev.2008.08.001

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070. doi:10.1111/1467-8624.00334

Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, *23*, 278–290. doi:10.1016/j.cogdev.2008.01.003

Siegal, M., & Peterson, C. C. (1996). Breaking the mold: A fresh look at children's understanding of questions about lies and mistakes. *Developmental Psychology*, *32*, 322–334. doi:10.1037/0012-1649.32.2.322

Siegal, M., & Peterson, C. C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology*, *34*, 332–341. doi:10.1037/0012-1649.34.2.332

Stipek, D. J., & Daniels, D. H. (1990). Children's use of dispositional attributions in predicting the performance and behavior of classmates. *Journal of Applied Developmental Psychology*, *11*, 13–28. doi:10.1016/0193-3973(90)90029-J

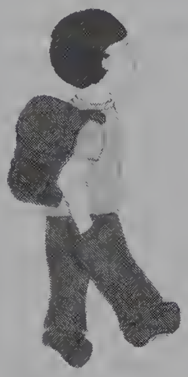
Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon says." *Child Development*, *44*, 849–853.

Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The development of distrust. *Child Development*, *82*, 1372–1380. doi:10.1111/j.1467-8624.2011.01629.x

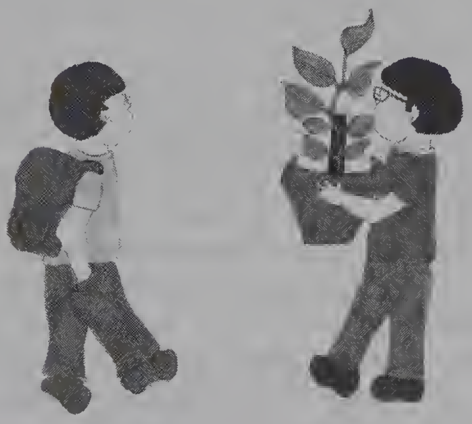
Appendix

Informant Introductions

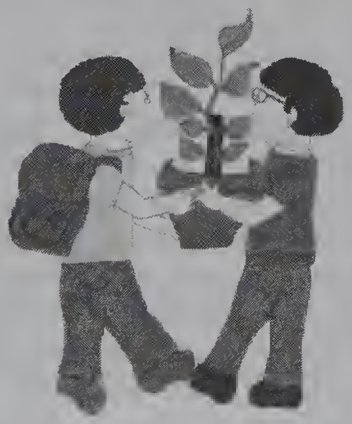
Nice



This is Neal. Neal is a boy who does nice things for people. Neal is nice.
Here's Neal at school. He is walking down the hall.

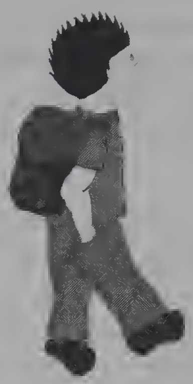


He walks by a boy carrying a plant.

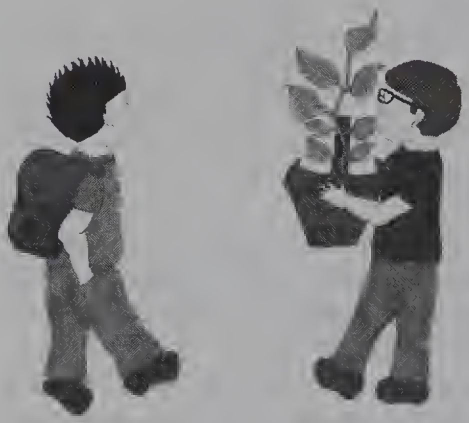


Look, Neal helps the boy carry the plant.
Remember, Neal is nice.

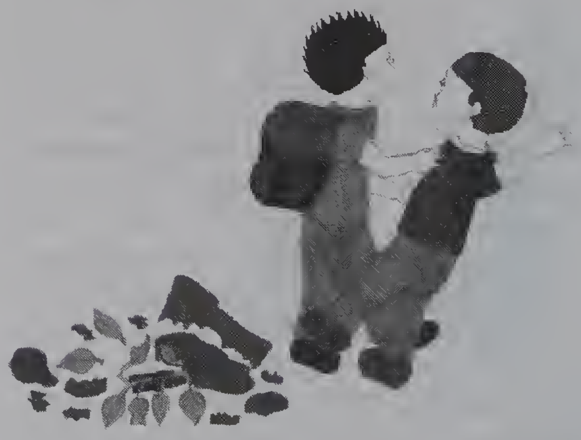
Mean



This is Mike. Mike is a boy who does mean things to people. Mike is mean.
Here's Mike at school. He is walking down the hall.

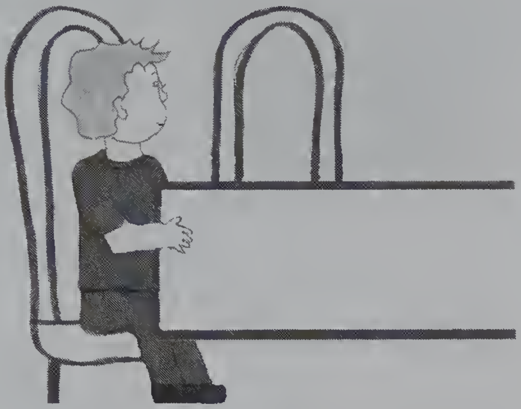


He walks by a boy carrying a plant.

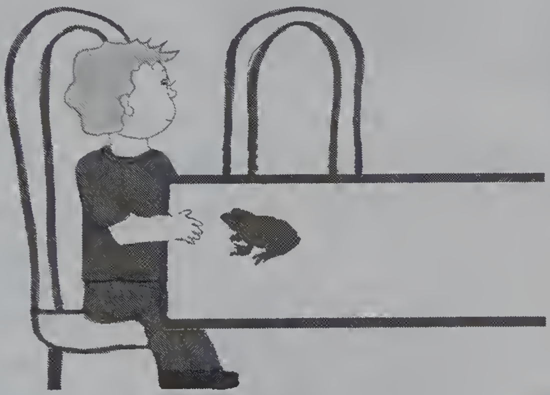


Look, Mike pushes the boy and his plant falls on the floor and breaks.
Remember, Mike is mean.

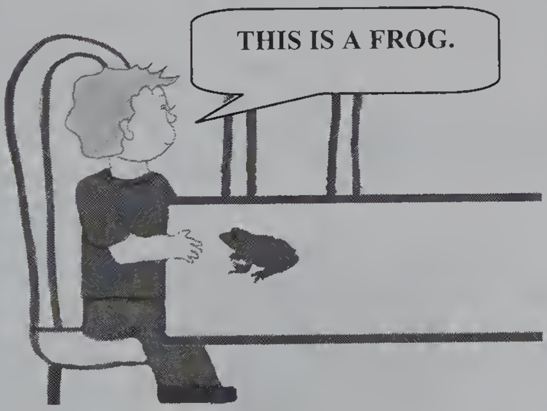
Smart



This is Sam. Sam is a boy who gives right answers. Sam is smart.

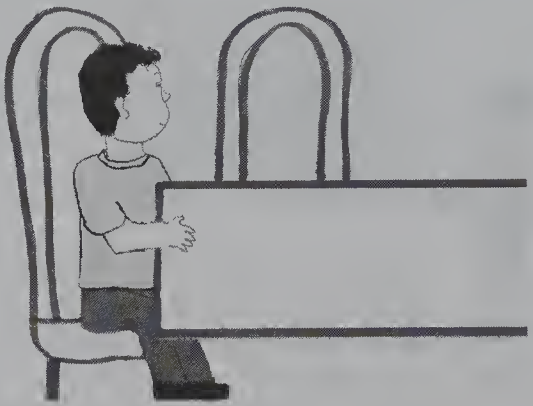


Here's Sam sitting at a table. Look at what's on the table.

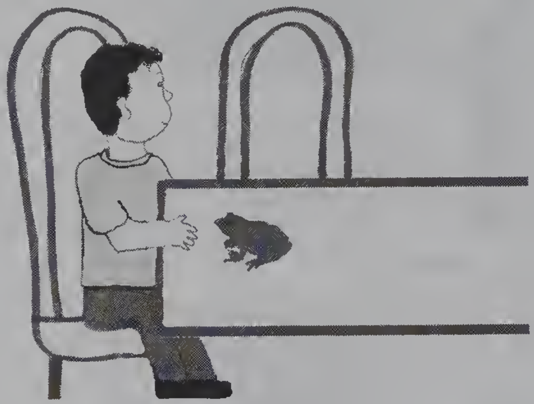

THIS IS A FROG.

Sam says: "This is a frog."
Remember, Sam is smart.

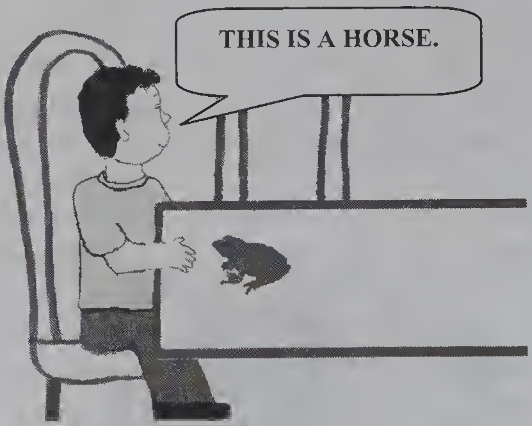
Not Smart



This is Noah. Noah is a boy who gives wrong answers. Noah is not smart.

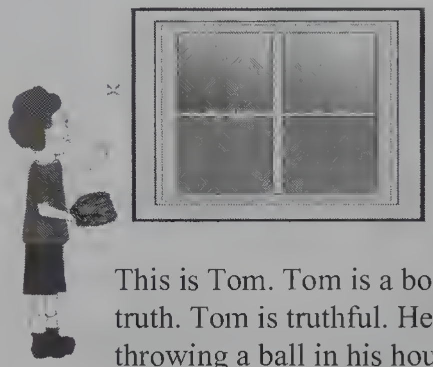


Here's Noah sitting at a table. Look at what's on the table.

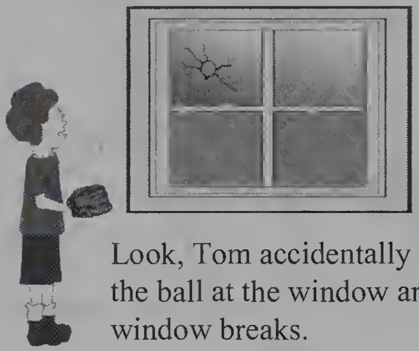

THIS IS A HORSE.

Noah says: "This is a horse."
Remember, Noah is not smart.

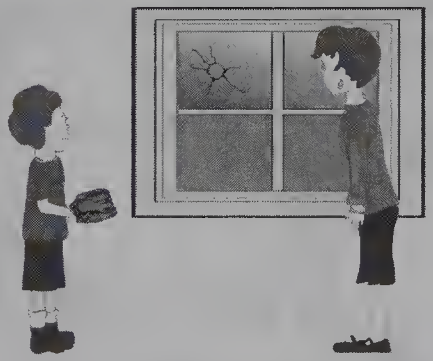
Honest



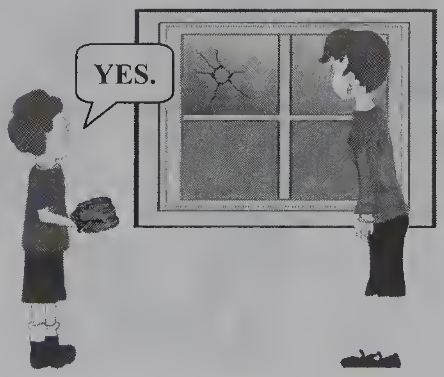
This is Tom. Tom is a boy who tells the truth. Tom is truthful. Here's Tom throwing a ball in his house.



Look, Tom accidentally throws the ball at the window and the window breaks.

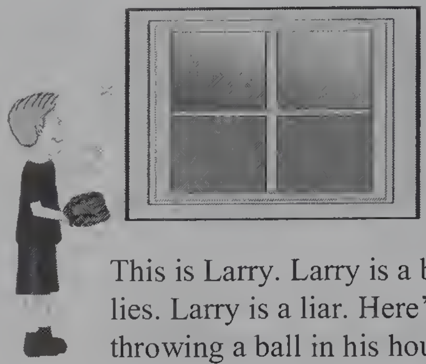


Tom's Mom comes in the room and asks, "Tom, did you break the window?"

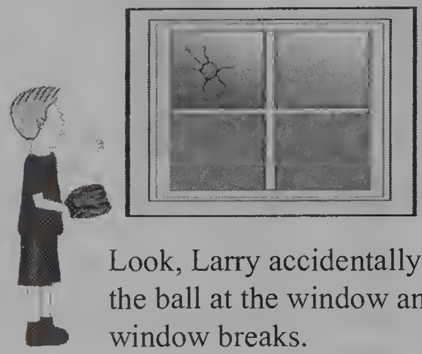


Tom says, "Yes." Remember, Tom is truthful.

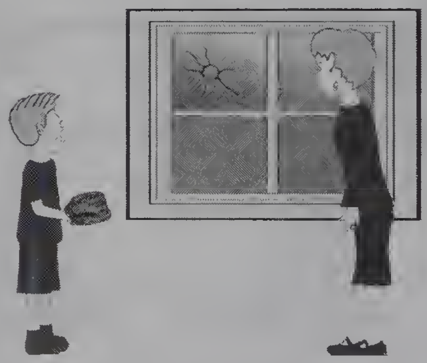
Dishonest



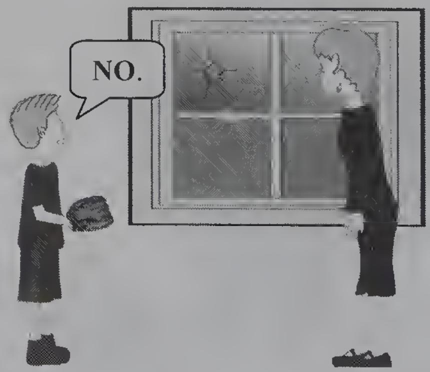
This is Larry. Larry is a boy who tells lies. Larry is a liar. Here's Larry throwing a ball in his house.



Look, Larry accidentally throws the ball at the window and the window breaks.



Larry's Mom comes in the room and asks, "Larry, did you break the window?"



Larry says, "No." Remember, Larry is a liar.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1269–1284

Rigidity in Gender-Typed Behaviors in Early Childhood: A Longitudinal Study of Ethnic Minority Children

May Ling Halim
California State University, Long Beach

Diane Ruble, Catherine Tamis-LeMonda,
and Patrick E. Shrout
New York University

A key prediction of cognitive theories of gender development concerns developmental trajectories in the relative strength or rigidity of gender typing. To examine these trajectories in early childhood, 229 children (African American, Mexican American, and Dominican American) were followed annually from age 3 to 5 years, and gender-stereotypical appearance, dress-up play, toy play, and sex segregation were examined. High gender-typing was found across ethnic groups, and most behaviors increased in rigidity, especially from age 3 to 4 years. In addressing controversy surrounding the stability and structure of gender-typing it was found that from year to year, most behaviors showed moderately stable individual differences. Behaviors were uncorrelated within age but showed more concordance in change across time, suggesting that aspects of gender-typing are multidimensional, but still show coherence.

A glance around a playground, a classroom, or a toy store will instantly reveal that early childhood is an important period in gender development for American children. Gender is displayed in highly visible, explicit ways. Flocks of girls, especially at ages 3 and 4 years, go about their days swathed in pink, glitter, hearts, and ribbons (Halim et al., 2013; Ruble, Lurye, & Zosuls, 2007). Boys, too, exhibit their own kind of "gender rigidity" (defined in this article as a strong adherence to gender-stereotypical behaviors; Parsons & Howe, 2006). Often, one can see boys kicking the air, flexing their muscles, and running around imagining themselves as Spiderman. This period when children display their gender so openly and ostentatiously provides a fascinating window into the early stages of gender development in the life span. The degree to which children espouse gender early on could set in motion life-long individual differences in displays of masculinity or femininity (Martin & Ruble, 2009).

Although the existence of obvious differences in gender-typed behaviors among young American boys and girls is clear, fundamental questions remain unanswered. First, we know little of whether different aspects of gender-typing emerge at the same time and connect to each other, or instead, exhibit specificity in expression. Second, questions remain about whether gender-typed behaviors show a predictable pattern of normative developmental change and stable individual differences over time or are a function of unpredictable situational variation. Third, it remains to be seen whether the above portrait of gender-typing can be found across different ethnic groups, or whether the focus on White, middle-class samples has led to distorted conclusions. Despite the explicit gender-typing seen in young children, there has been little empirical investigation of it—over time, across different behaviors, and across different populations. Some may explain these behaviors as just a Disney-provoked middle-class display. However, we argue that these behaviors reflect a manifestation of a fundamental developmental phenomenon that marks the early phases of

---

This work was conducted at New York University's Center for Research on Culture, Development, and Education (CRCDE), within the Department of Applied Psychology at New York University's Steinhardt School of Culture, Education, and Human Development. This research was supported by National Science Foundation Grants BCS 021859 and IRADS 0721383 to C. S. Tamis-LeMonda. Preparation of this paper was also supported in part by a National Institute of Child Health and Human Development Research grant (R01 HD04994 and ARRA Supplement) to D. N. Ruble. We thank our colleagues and staff at the CRCDE, particularly Cristina Hunter, Eva Liang, Yana Kuchirko, Julia Raufman, Emerald Shee, Irene Sze, and Irene Wu, as well as the mothers and children who participated in our research. We also thank David Amodio and Kay Deaux for their feedback on earlier versions of this article.

Correspondence concerning this article should be addressed to May Ling Halim, California State University, Department of Psychology, 1250 Bellflower Blvd., MS0901, Long Beach, CA 90840. Electronic mail may be sent to mayling.halim@csulb.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0012
DOI: 10.1111/cdev.12057

a life course of a gender divide. Thus, investigating these questions is critical in adding to existing literature on gender development.

These questions are unanswered because the literature to date is primarily composed of cross-sectional studies of middle-class White children. Exciting new insights can be obtained by broadening the study of gender development to include samples of ethnically diverse children in a longitudinal design. Insofar as children raised in the United States experience the same global culture during the developmentally important preschool years, and insofar as increasing rigidity in gender-typed behaviors reflect a fundamental, generalizable developmental theory, they should show the same patterns of gender development. This study aims to address these questions by examining the change and stability (consistency of a behavior across time) of different facets of gender-typing in an ethnically diverse sample of 3- to 5-year-old children.

Trajectories of Gender Development in Early Childhood

According to cognitive theories of gender development, as children learn about gender categories, they are predicted to be highly motivated to strictly adhere to gender stereotypes, and this adherence may increase in rigidity as information about gender categories is constructed and consolidated (Martin, Ruble, & Szkrybalo, 2002; Ruble, 1994). These theories view children as active "gender detectives" (Martin & Ruble, 2004), who earnestly seek out information about gender, then attempt to draw inferences and clear conclusions about gender, and finally apply such conclusions rigidly to themselves. Subsequently, cognitive theories of gender development predict that children will integrate their new gender identity with other identities and become more flexible in the application of the conclusions they drew about gender. These theories are, of course, not the only perspectives on gender development. Socialization approaches posit that experiences at home, school, and one's neighborhood influence children's gender-typing through processes such as social learning and encouragement from others to behave in gender-typed ways. Biological approaches emphasize the role of genes, hormones, and the brain in physical and psychological sexual differentiation (Ruble, Martin, & Berenbaum, 2006). These perspectives on gender development would predict a high level of gender-typed behaviors in early childhood. However, only cognitive theories of gender development take an explicitly developmental perspective, emphasizing not just high levels, but change over time, upon which we base this study.

In this study we employed a longitudinal design to investigate developmental growth within the same individuals. We asked whether gender-typed behaviors *increase* in rigidity during ages 3 to 5 years, as this period is a time of increasing gender knowledge, rather than just whether rigidity is high. We also explored whether this rigidity begins to level off during these ages; that is, do some gender-typed behaviors show the curvilinear pattern predicted by cognitive theories of gender development?

This question is important because most studies on the developmental trajectory of gender-typing focus primarily on gender-related beliefs and cognitions. Several studies have shown that children's gender stereotyping increases in rigidity from ages 3 to 5 years, which is then generally followed by flexibility (Signorella, Bigler, & Liben, 1993; Trautner et al., 2005). Other research has found similar increases in rigidity from ages 3 to 4 years in attitudes about children who violate gender norms (Ruble et al., 2007). But how do manifestations of rigidity in gender-related *behavior* develop across early childhood? Do different types of gendered behaviors show a similar course of increasing rigidity as seen in gender-related cognitions? Past research examining developmental trajectories in gender-typed behavior and preferences has largely been limited to either peer or toy preferences. This research, largely cross-sectional with a few longitudinal studies, has found that children increasingly prefer to play with same-sex children from preschool through the elementary school years (e.g., Maccoby & Jacklin, 1987; Martin, Fabes, Evans, & Wyman, 1999; Serbin, Powlishta, & Gulko, 1993) and display increased rigidity in gender-typed toy preferences (e.g., Servin, Bohlin, & Berlin, 1999).

In summary, there is little research on the trajectories of multiple aspects of gender-typed behavior during the preschool years, as well as little attempt to pull together the different facets of gender-typing in a single study. This study addresses this gap in the literature by looking at four different behaviors within one study over time. We were interested in how children adorn themselves as it is relatively an unexplored element of gender development and seems both prevalent among children and important to them. Thus, we observed children's gender-typed appearances and their involvement in dress-up play. Gender researchers more commonly study activities and peer choice, so we included these elements in this study to provide reference points for comparison. On the basis of the aforementioned studies and in line with proposals that gender-typing is multidimensional (Huston, 1983; Ruble et al., 2006), we anticipated different patterns of normative change for different types of behaviors. Physical appearances are salient to children at a very early age (Ruble, Lurye, et al., 2007), and thus the process of construction and consolidation might "take off" between ages 3 and 4 years (Ruble, 1994). Consistent with this idea, in a study of White, middle-class girls, appearances were the most gender-typed at ages 3 and 4 years and lower at ages 5 and 6 years (Halim et al., 2012). For these reasons, we predicted that children's actual gender-typed appearances and their engagement in dress-up play would be characterized by a curvilinear pattern across ages 3–5 years with gender stereotypicality increasing to a peak and then leveling off or declining. In contrast, participation in gender-typed play activities and sex segregation were expected to continue to increase from ages 3 to 5 years.

Multidimensional Nature of Gender-Typing

The gender development field has conceptualized gender-typing as a multidimensional matrix (Huston, 1983; Ruble et al., 2006). Is there unity in gender-typing? Or does specificity better describe the way children express their gender? The longitudinal design of this study allows us to look at the coherence of gender-typing both within and across time. On one hand, there has been a prevalent implicit assumption of coherence in the literature (e.g., if a girl likes to play with dolls, she probably loves pink as well). Indeed, the very construct of gender-typing assumes that there are factors that lead children to become gender-typed as they acquire a set of behaviors and interests that are more typical of their own gender than of the other gender (Maccoby, 2002). Thus, if one is to study gender-typing, there needs to be something coherent to study, even if coherence occurs only across a few attributes, a view evident in many prominent traditions (e.g., Martin & Halverson, 1981). On the other hand, specificity in gender manifestations is a possibility; there have been many failures to find relations among all types of gender-related variables (Ruble & Martin, 1998). There are also studies that find no correlation between gender-typed preferences or behaviors and gender-stereotype knowledge. Unity may be more evident in smaller subsets of gender-typing variables (e.g., only within children's gender attitudes; Ruble & Martin, 1998). Looking at cohesion both within time and across time may shed light on this controversy. On the basis of past research (Martin & Fabes, 2001), we expected that gender-typed play and sex segregation might correlate with one another both within and across time, as these two elements feed into each other bidirectionally.

Stability in Gender-Typing

Another vital and controversial question is whether or not individual differences in gender-typing are stable over time even in the context of normative changes of increasing gender-typed behaviors (Martin & Ruble, 2009). Does a girl who is highly gender-typed compared to her peers at age 3 years remain so at age 4 years? Are certain aspects of gender-typing (e.g., play activities) more stable than others (e.g., sex segregation) across the ages of 3, 4, and 5 years? These questions are important for understanding the processes involved in gender development over the life span and could have implications for self-perceptions and life choices, such as the valuing of academic domains and occupational aspirations, in ways that would be stronger than mere fleeting gender proclivities. Yet, surprisingly, little empirical data exist on this topic, and the question of the stability of individual differences in gender-typing is an area of contention.

Prominent gender development theories would predict stability in gender-typing, if even for a short time, as hypothesized predictors of gender-typing, such as hormones or parent modeling of gender-typed behavior, also tend to be stable in childhood (see Golombok, 2008). However, some gender researchers have concluded that, actually, the limited data available show little stability in gender-typing (Maccoby, 2002). Instead, Maccoby (2002) suggested that ample variation in situations surrounding a child affects how gender-typed a child may appear. Furthermore, any stability that is found may reflect group-level associations (e.g., classroom peer influences) rather than individual-level ones. Other psychologists have challenged this viewpoint (Martin & Ruble, 2009). Some studies indicate moderate stability in sex segregation in preschool to elementary school-aged children under some conditions (e.g., Martin & Fabes, 2001; Serbin et al., 1993), and other studies indicate moderate stability in gender-typed play and toy preferences among preschoolers and kindergarteners (Golombok et al., 2008; Maccoby & Jacklin, 1974;

Martin & Fabes, 2001). In this study, we sought to address this controversy by examining several behaviors in a single study to determine whether certain aspects of gender-typing show more stability than others.

Rigidity in Gender-Typed Behaviors Across Ethnic Groups

In this study we also aimed to test the robustness of the phenomenon of increasing gender-typed behaviors across ethnic groups, as the vast majority of gender development research has only sampled White middle- to upper-middle-class children. Because of this bias, only tentative statements as to what is normative in gender development can be made. We examined African American and Latino immigrant children because these ethnic groups widely differ on factors that could affect gender-typing, which would allow us the most stringent test of the generalizability. In addition, these groups have an important presence in the United States—African Americans make up a substantial minority, and the population of Latinos is increasing (U.S. Census Bureau, 2011). Because we consider increasing gender-typed behaviors to be a fundamental developmental phenomenon driven largely by changes in cognition, we expected curvilinear trajectories across all ethnic groups. However, we explored ethnic variation in overall levels or in the timing of those trajectories based on tentative evidence showing differences in gender roles and gender socialization by ethnic group. For example, scholars have argued that African American women's ample participation in the workforce has brought about more egalitarian gender roles and less sharply differentiated gender stereotypes in African American families (e.g., Stanback, 1985), which has received some empirical support (Jarret, Roy, & Burton, 2002). Also, African American 4- to 6-year-old children have shown less gender stereotype endorsement than White children (Albert & Porter, 1988), although this has not always been replicated (Liben & Bigler, 2002). In contrast, *machismo* and *marianismo*—or, broadly, male dominance and female submissiveness—are key constructs in Latino culture (DeSouza, Baldwin, Koller, & Narvaz, 2004). Further, Latino children have been found to endorse occupational gender stereotypes more than White American children (Bailey & Nihlen, 1990). And a higher proportion of Mexican and Dominican immigrant 2-year-olds were able to produce gender labels compared to African American 2-year-olds (Zosuls, 2008). On the basis of this sparse literature, we explored whether Latino children might exhibit higher levels of gender-typed behavior and show peak gender rigidity earlier compared to African American children.

Study Overview

This study assessed four gender-typed behaviors in African American, Mexican, and Dominican children from a large urban Northeastern city. At ages 3, 4, and 5 years, coders rated how gender-typed the child's appearance was and gathered information from mothers on children's participation in dress-up play, participation in gender-typed play activities, and the number of girls and boys that were children's friends. We examined three types of hypotheses: (a) specific theory-driven hypotheses, (b) those addressing unresolved controversies, and (c) exploratory hypotheses. For the first type, on the basis of cognitive theories of gender development, we predicted that across gender and ethnic groups children would show high levels and curvilinear trajectories of increasing rigidity in gender-typed behaviors across ages 3–5 years. Furthermore, we hypothesized that appearance-related behaviors would peak earlier in rigidity and show some leveling off compared to other behaviors. For the second type, we investigated whether these different behaviors would be connected to each other in a constellation of gender-typing and whether they paralleled one another in how they changed. We also investigated whether, in the middle of normative changes, there would be stable individual differences in gender-typing. Finally, for the third type, we explored whether the levels or trajectories of these behaviors would show variation by ethnic group.

Method

Participants and Procedure

Preschool-aged children ($N = 229$; 108 girls, 121 boys) and their mothers ($M_{\text{age}} = 29.06$, $SD = 5.64$) participated: 84 Dominican American (49 boys, 35 girls), 77 African American (42 boys, 35 girls), and 68 Mexican American (30 boys, 38 girls). We assessed preschoolers at three waves: at age 3 years ($M = 3.03$ years, $SD = 0.14$), 4 years ($M = 4.21$ years, $SD = 0.16$), and 5 years ($M = 5.16$ years, $SD = 0.12$). Of the 229 families, 154 families participated at all three waves, 28 families participated at two of the three waves, and 47 families participated at only one of the three waves. Comparing families with complete versus incomplete data revealed no significant differences in mother's education, work status, immigrant status, marital status, cohabitation status, dominant language, age, physical health, psychological distress, household income, or child cognitive or language skills at age 2. However, girls and Mexican children were more likely to have complete data.

Participants were recruited as part of a larger longitudinal study from three public hospitals in a large Northeastern urban city shortly after mothers delivered their children. Eligibility to participate in the study required the mothers to (a) be at least 18 years old; (b) not live in a shelter; (c) have a healthy, full-term infant (birth weight > 2,500 g); and (d) self-identify as Mexican, Dominican, or U.S.-born African American. Of all mothers approached who were eligible, 50% agreed to participate. Average family annual income was $20,500 ($SD$ = $14,751). Dominican mothers were 77% first generation, and 23% second generation, and were primarily from Santo Domingo. Mexican mothers were 96% first-generation immigrants and were primarily from Puebla. At the infants' births, 73% of Dominican, 48% of Mexican, and 65% of African American mothers completed at least a high school education or a GED. Father coresidency rates were highest in Mexican families (87%), followed by Dominican (63%) and African American (46%) families, $F(2, 216) = 14.44$, $p < .001$.

Interviews took place either at children's homes (age 3 years) or at a university (ages 4 and 5 years). We obtained parental consent either in person or through the mail via signed returned consent forms. We paid mothers $75.00 for each interview at each wave.

Measures

Observations of gender-typed appearance. We coded children's gender-typed appearance from videotapes of the interviews at each wave. Researchers coded whether they observed a set of specific gender-typed elements (0 = *not present*, 1 = *present*). For girls, elements included the following (κs range = .70–1.00): dresses or skirts, feminine colors (e.g., pink), feminine hair accessories (e.g., bows), feminine patterns or logos (e.g., flowers), feminine fabric or fit (e.g., taffeta), trend-conscious styles (e.g., knee-high boots), formal wear (e.g., patent-leather Mary Janes), and jewelry (e.g., bracelets). For boys, elements included the following: masculine colors (e.g., red and black together), masculine patterns or logos (e.g., graffiti), masculine fabric or fit (e.g., baggy pants), sports-themed styles (e.g., baseball jerseys), and formal wear (e.g., skinny tie with a vest). To create a score of gender-typed appearance we summed elements for girls and boys separately. Girls could reach a maximum score of 8 ($M = 3.29$, $SD = 1.66$) and boys could reach a maximum score of 5 ($M = 2.29$, $SD = 0.99$). Girls' and boys' scores had different possible ranges because few boys showed trend-conscious styles or jewelry, and most had short hairstyles. For comparison purposes, we created z scores for girls' and boys' scores, each centered around the grand mean (across ages 3, 4, and 5 years) within gender.

Gender-typed dress-up play. Interviewers asked mothers: "Over the past month, how often did your child play with dress-up clothes or costumes like princess or Spiderman costumes, pocketbooks, or adult-like shoes?" (0 = *never*, 1 = *once or twice a month*, 2 = *once a week*, 3 = *several times a week*, 4 = *everyday*; $M = 1.53$, $SD = 1.50$). Interviewers also asked mothers to list the type of dress-up clothes or costumes with which their child played. Only gender-typed costumes or clothes were counted (excluding five children at age 3 years, two boys at age 4 years, and one girl at age 5 years who liked to dress-up as Michael Jackson). The most common themes included dressing up as a princess for girls (especially Disney princesses at age 4 years) and dressing up as some sort of superhero for boys (especially at age 5 years). Other themes included dressing up in adult-like clothing, including the clothing of parents, siblings, or relatives (especially at age 3 years), and dressing up as some sort of TV or cartoon character (e.g., Hannah Montana or a Transformer).

Gender-typed play. Interviewers asked mothers how often over the past month their children played with three feminine items—(a) kitchen sets, tea sets, and/or food sets; (b) dolls; and (c) soft toys like teddy bears, stuffed animals, or puppets)—and three masculine items—(a) toy guns or objects used as guns and/or swords or objects used as swords; (b) vehicles like cars, trucks, and trains; and (c) balls (0 = *never*, 1 = *once or twice a month*, 2 = *once a week*, 3 = *several times a week*, 4 = *everyday*). These toys were chosen based on many studies using the same stimuli (Ruble & Martin, 1998). Prior research has shown that mother report of children's toy play has good reliability across time (Golombok et al., 2008). To make a scale, for girls we reverse coded the three masculine items and averaged them together with the three feminine items ($M = 2.74$, $SD = 0.54$). For boys we reverse coded the three feminine items and averaged them together with the three masculine items ($M = 2.70$,

$SD = 0.55$). We then combined girls' and boys' scores into a single measure ($\alpha_{age3} = .59$, $\alpha_{age4} = .70$, $\alpha_{age5} = .72$).

Sex segregation. Interviewers asked mothers: "Does your child have the opportunity to spend time with other children?" If mothers answered, "Yes," they were asked, "Who are they?" Mothers could list up to nine of their child's peers and reported the peer's sex, the peer's age, how the peer was related to the child, and if the peer was actually the child's friend. To be consistent with prior studies on sex segregation, we excluded peers that were either over age 10 years or siblings and only included peers who were counted as friends. At each age we calculated the proportion of same-sex peers of total peers named ($M = 63.01\%$, $SD = 30.00\%$).

Results

We first discuss findings pertaining to normative changes in gender-typed behaviors from ages 3 to 5 years and report any cultural or gender variation in the level and trajectory of these behaviors. Next we report whether these different behaviors are connected to each other in a constellation of gender-typing by examining their associations both (a) within time and (b) across time as the behaviors change. Finally, we report the stability of gender-typing across consecutive years.

Change in Gender-Typed Behavior Over Time and Group Differences

Analytic strategy. We examined change over time of gender-typed behavior by conducting multilevel models for each behavior, which included children with at least two waves of data. For each model, we first calculated Type III sums of squares to indicate overall omnibus fixed effects for time (centered at age 4 years), gender (female reference group), ethnicity (Dominican reference group), and interactions among the three variables and a random participant intercept effect in predicting each behavior. For ease of communication we present the results using fitted means from the mixed models. For detailed model coefficients and formulas, see Halim (2012). We treated gender and ethnicity as factors and time as a covariate, and we assumed that the residuals were uncorrelated and homoskedastic over time. Next, we repeated the model with the linear and quadratic components of time separated. For follow-up analyses, we repeated the mixed models, but with alterations (e.g., treated time as a factor for post hoc time period comparisons; used effects coding to look at effects across gender or across ethnicity).

Observations of gender-typed appearance. Overall, children, especially girls, were very gender-typed in their appearances. Dominican girls were more gender-typed in their appearance compared to Mexican girls. Furthermore, as predicted, gender-typed appearance followed a curvilinear pattern across time for the sample as a whole. Trajectories varied across ethnic groups. See Figure 1a and Table 1.

Specifically, the mean number of feminine appearance elements at age 4 years was 3.73 ($SD = 1.97$). Although a mean of 3.73 of a possible 8 may seem low, qualitatively, it indicates that on average, girls exhibited between 3 and 4 gender-typed elements, such as wearing a pink dress with hearts and an additional big bow in her hair, creating a very feminine appearance. Boys exhibited between 2 and 3 elements of 5 overall at each age, indicating overall masculine appearances as well,

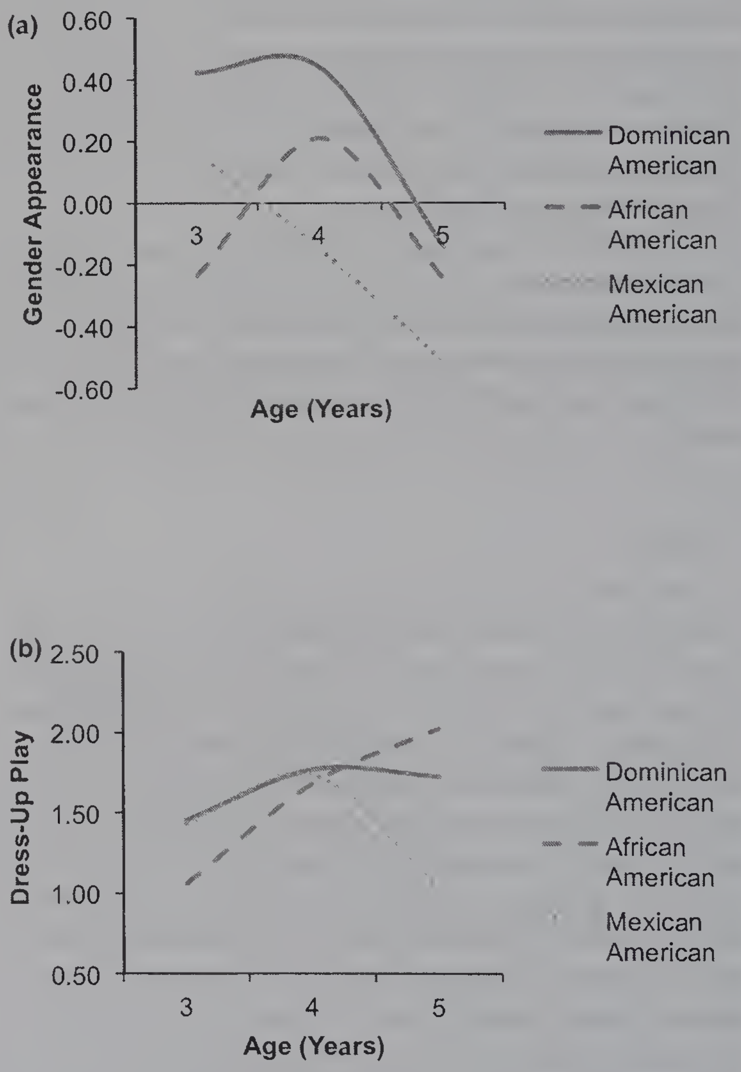


Figure 1. (a: top panel) Means of observed gender-typed appearance over time by ethnic group. (b: bottom panel) Means of gender-typed dress-up play frequency over time by ethnic group.

Table 1

Means and Standard Deviations of Gender-Typed Behaviors

	Age 3 years		Age 4 years		Age 5 years		Across time	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Observed appearance rigidity								
Boys overall	2.52	0.94	2.36	1.05	1.99	0.91	2.28	0.99
Girls overall	$3.32_{a,b}$	1.32	3.73_{a}	1.97	2.80_{b}	1.51	3.28	1.66
African American children[a]	$-0.24_{a,d}$	0.68	$0.23_{b,d,e}$	1.04	$-0.24_{a,d,e}$	0.93	-0.07_{d}	0.93
Mexican American children[a]	$0.11_{a,e}$	0.96	$-0.15_{a,b,d}$	1.05	$-0.52_{b,d}$	0.86	-0.18_{d}	0.99
Dominican American children[a]	$0.43_{a,e}$	0.85	$0.37_{a,e}$	1.23	$-0.16^{b,e}$	0.93	0.20_{e}	1.05
All children overall[a]	0.12_{a}	0.88	0.16_{a}	1.13	-0.30_{b}	0.91	−0.01	1.00
Gender-typed dress-up play								
Boys overall	$0.71_{a,d}$	1.13	$1.29_{b,d}$	1.35	$1.22_{a,b,d}$	1.38	1.08_{d}	1.31
Girls overall	1.89_{e}	1.6	2.15_{e}	1.57	1.96_{e}	1.50	2.00_{e}	1.55
African American children	1.02_{a}	1.45	$1.63_{a,b}$	1.54	$1.96_{b,d}$	1.49	1.56	1.54
Mexican American children	$1.51_{a,b}$	1.55	1.76_{a}	1.60	$1.08_{b,e}$	1.29	1.46	1.50
Dominican American children	1.31	1.47	1.71	1.43	$1.63_{d,e}$	1.53	1.55	1.48
All children overall	1.29_{a}	1.49	1.70_{b}	1.52	$1.57_{a,b}$	1.48	1.52	1.50
Gender-typed play								
Boys overall	2.60_{a}	0.56	$2.73_{a,b}$	0.62	2.78_{b}	0.46	2.70	0.55
Girls overall	2.64_{a}	0.53	$2.74_{a,b}$	0.54	2.81_{b}	0.50	2.73	0.53
African American children	2.52_{a}	0.61	$2.76_{b,d,e}$	0.59	$2.68_{a,b,d}$	0.52	2.66	0.58
Mexican American children	2.64_{a}	0.55	$2.61_{a,b,d}$	0.65	$2.80_{b,d,e}$	0.52	2.66	0.58
Dominican American children	2.68_{a}	0.48	$2.82_{a,b,e}$	0.50	$2.89_{b,e}$	0.38	2.80	0.46
All children overall	2.62_{a}	0.55	2.73_{b}	0.59	2.80_{b}	0.48	2.72	0.54
Sex segregation								
Boys overall	0.48_{a}	0.29	$0.71_{a,b}$	0.26	0.73_{b}	0.28	0.65	0.30
Girls overall	0.49_{a}	0.28	$0.68_{a,b}$	0.29	0.75_{b}	0.25	0.64	0.30
African American children	0.55	0.28	0.66	0.27	0.71	0.30	0.65	0.29
Mexican American children	0.44_{a}	0.28	0.70_{b}	0.32	0.78_{b}	0.25	0.64	0.32
Dominican American children	0.46_{a}	0.29	0.72_{b}	0.25	0.73_{b}	0.25	0.64	0.29
All children overall	0.48_{a}	0.29	0.70_{b}	0.28	0.74_{b}	0.27	0.64	0.30

Note. Means not sharing the same (a) a, b, c subscripts within row, and (b) the same d, e subscripts within columns and sections are significantly different at $p < .05$. Subscripts are only given where findings were significant.

[a] Means given for boys and girls separately are in original scale for ease of interpretability. Means when girls and boys are combined are the *z*-scored means.

such as wearing dark, baggy clothes with a graffiti pattern.

In exploring ethnic differences we found that across time, the model revealed a main effect of ethnicity, $F(2, 485.4) = 5.49$, $p = .004$, qualified by a Gender × Ethnicity interaction, $F(2, 485.4) = 6.79$, $p = .001$, such that the effect was concentrated in the girls. Dominican girls ($M = 4.05$, $SD = 1.74$) were significantly more gender-typed in their appearances compared to Mexican girls ($M = 2.82$, $SD = 1.58$), $p = .015$, but not compared to African American girls ($M = 3.04$, $SD = 1.37$), $p = .165$. Among boys, there were no differences in gender-typed appearance among the three ethnic groups.

The average level of gender-typed appearance changed over time, and the overall effect was curvilinear as expected: time linear, $F(1, 337.1) = 16.40$, $p < .001$, and time quadratic, $F(1, 330.9) = 8.65$, $p = .004$. The level of gender-typed appearance was not different from ages 3 to 4 years ($p = .646$). Age 5 gender-typed appearance was lower than age 3 and 4 levels (*ps* < .001). We also found an interaction between time and ethnicity, $F(2, 333.8) = 2.83$, $p = .025$. Both Dominican (time linear; $p = .001$; time quadratic: $p = .041$) and African American children showed significant curvilinear patterns (time linear: *ns*; time quadratic: $p = .004$). Mexican children showed a significant negative linear trend over time ($p < .001$). See Figure 1a. The interaction between time and gender was not significant, $F(2, 333.9) = 1.98$, *ns*.

Gender-typed dress-up play. As predicted, girls played dress-up more frequently than boys at all ages, $F(1, 175.2) = 35.36$, $p < .001$. We explored whether Latino children played dress-up more frequently than African American children, but there were no differences among the three ethnic groups, $F(2, 175.3) = .91$, *ns*, nor any Gender × Ethnicity interaction, $F(2, 175.3) = 1.21$, *ns*. Furthermore, as expected, dress-up play followed a curvilinear pattern across time and we again found cultural variation in these trajectories. Specifically, the overall dress-up play mean of 2.00 for girls can be interpreted as girls playing dress-up about once a week on average, whereas the boys' mean of 1.08 can be interpreted as playing dress-up about once or twice a month on average during the preschool years. Examined another way, 56% of girls and 31% of boys played dress-up anywhere from once a week to every day across the preschool years.

As for developmental change, we found a significant main effect for time, $F(2, 325.3) = 5.26$, $p = .006$, with both linear, $F(1, 327.0) = 4.58$, $p = .033$, and quadratic, $F(1, 323.7) = 5.99$, $p = .015$, components. Dress-up play increased in frequency from ages 3 to 4 years ($p = .002$), then leveled off in frequency from ages 4 to 5 years, but this small decrease was not significant ($p = .246$). The model also revealed an interaction between time and ethnicity, $F(4, 325.3) = 4.72$, $p = .001$. African American children increased linearly over time in dress-up play frequency ($p < .001$). Mexican children increased then decreased in dress-up play frequency, peaking at age 4 years (quadratic: $p = .011$). Dominican children did not significantly change over time. See Figure 1b. The interaction between time and gender was not significant, $F(2, 325.3) = .76$, *ns*.

Gender-typed toy play. We first conducted mixed models on the full six-item gender-typed play scale followed by exploration of a breakdown of the scale: three items of cross-gender-typed play and three items of same-gender-typed play. Both boys and girls showed high levels of same-gender-typed play (approximately once a week to several times a week) and low levels of cross-gender-typed play (approximately one to two times a month), and boys and girls did not significantly differ in their levels, $F(1, 175.9) = .27$, *ns*. An exploration of differences by ethnic group, $F(2, 176.0) = 2.56$, *ns*, or by gender and ethnic group, $F(2, 176.0) = 2.33$, *ns*, revealed no significant effects.

Gender-typed toy play (six-item scale) linearly increased over time across ethnicity and gender, $F(2, 328.1) = 6.67$, $p = .001$. Children engaged in more gender-typed play at ages 4 and 5 years compared to age 3 years, $p = .010$ and $p < .001$, respectively. This pattern was driven by linearly decreasing levels of cross-gender-typed play, $p < .001$ ($M_{\text{age3}} = 1.40$, $SD_{\text{age3}} = .84$; $M_{\text{age4}} = 1.22$, $SD_{\text{age4}} = .89$; $M_{\text{age5}} = .97$, $SD_{\text{age5}} = .82$). Girls' female-typed and boys' male-typed play was high at all time points and did not change.

Sex segregation. At age 3 years, 48% of children's friends were of the same gender; at age 4 years, 70% were of the same gender; and at age 5 years, 74% were of the same gender. As predicted, sex segregation increased over time in a curvilinear fashion, and there was no evident variation by ethnicity, $F(2, 164.8) = .01$, *ns*; or gender, $F(1, 164.8) = .05$, *ns*; or interaction between the two, $F(2, 164.8) = 1.12$, *ns*. Specifically, the model revealed a main effect for time, $F(2, 310.9) = 39.04$, $p < .001$; linear, $F(1, 312.2) = 69.02$, $p < .001$; and quadratic, $F(1, 310.0) = 10.10$, $p = .002$. Sex segregation increased from ages 3 to 5 years with the sharpest increase from ages 3 to 4 years (from 3 to 4 years, $p < .001$; from 4 to 5 years, $p = .094$).

Summary. As expected, children exhibited rigid male- and female-typed behaviors during the early childhood years. They dressed and played in gender-typed ways and mainly were friends with same-gender peers. We found few ethnic differences in overall levels of gender-typing except that Dominican girls dressed in more stereotypical ways than did Mexican girls.

Our expectations for the developmental pattern of increased rigidity in gender-typed behaviors were supported as well. From ages 3 to 5 years children's friends were increasingly same-gender peers. Also, children increasingly avoided cross-gender-typed play. Showing a slightly different pattern, children increasingly played dress-up from ages 3 to 4 years then remained steady. In contrast, children's gender-typed appearances showed some movement toward flexibility, remaining highly stereotypical at ages 3 and 4 years, then becoming less so from ages 4 to 5 years. Our exploration of ethnic variation in trajectories showed ethnic differences for two of the four behaviors (appearance and dress-up play).

Coherence in Gender-Typed Behaviors

The second goal was to examine the multidimensional nature of gender-typing and speak to the controversy surrounding whether or not different aspects of gender-typing develop in concert. We first tested whether exhibiting gender rigidity in one behavior tended to be associated with rigidity in other behaviors at each age by examining correlations for the sample as a whole, by gender, and by ethnicity, similar to what most of the little exist-

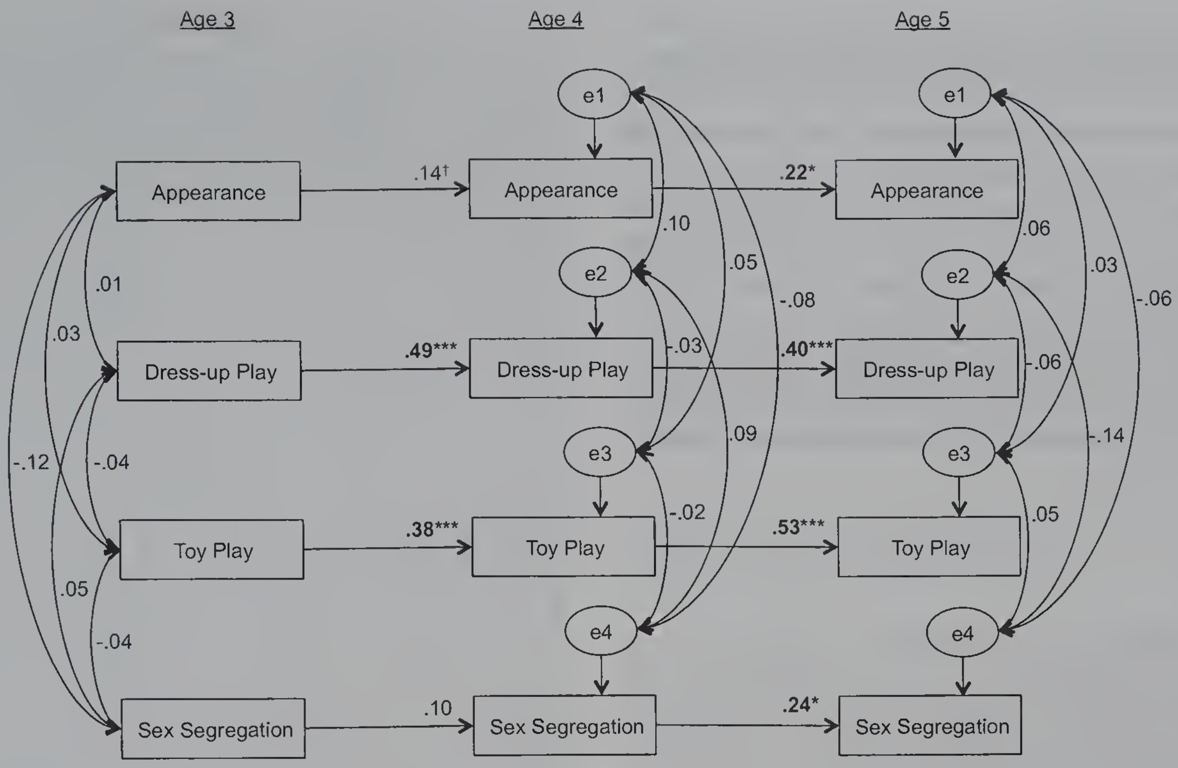


Figure 2. Structural equation model showing within-age correlations and stability estimates among gender-typed behaviors. Standardized coefficients are depicted. Coefficients not in boldface are nonsignificant.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{***}p < .001$.

ing research has done (Martin & Ruble, 2009). We next expanded upon this literature and tested whether a change in one gender-typed behavior from one year to the next was related to change in other gender-typed behaviors across the same time period. The focus on change has received new attention in recent years (McArdle, 2009). On the basis of the assumption that a coherent developmental process underlies the changes, we hypothesized that the various behaviors would be moderately associated with each other, within and across age. Given the number of associations we wanted to investigate, and the possibility of making Type I errors, to estimate within-age associations we conducted a structural equation model that simultaneously included all associations. We used Mplus to estimate the model and used full information maximum likelihood estimation, which allowed participants with incomplete data to be included. We also estimated the across-age correlations in the context of Mplus because it allowed the matrix of correlations to be estimated simultaneously and also allowed the handling of missing data under the assumption of a data missing at random pattern.

Within-age correlations among gender-typed behaviors. In the model, we tested associations among different gender-typed behaviors within each age ($N = 145$; Figure 2). To investigate stability of gender-typing within measures, we simultaneously tested whether gender-typing scores at one age predicted gender-typing scores at the consecutive age, which will be discussed later in this section. Results indicated satisfactory model fit, $\chi^2(40) = 55.63$, comparative fit index (CFI) = .887, root mean square error of approximation (RMSEA) = .052. Unexpectedly no predicted associations were found between any measures at any age (Figure 2). Furthermore, multiple groups analyses indicated no significant improvement in model fit when separating the model by gender, $\chi^2(8) = 12.57$, *ns*, or ethnicity, $\chi^2(16) = 24.02$, *ns*.

Change-to-change associations among gender-typed behaviors. We calculated difference scores for the four gender-typed behaviors (e.g., dress-up play at age 3 years from dress-up play at age 4 years) and assessed associations among these difference scores. Three of the 12 correlations were significant (Table 2). An increase in sex segregation was positively associated with an increase in dress-up play from age 3 to 4 years, $r(176) = .17$, $p = .024$, and with an increase in gender-typed play from 4 to 5 years, $r(176) = .16$, $p = .034$. Unexpectedly, an increase in sex segregation from 4 to 5 years was associated with a decrease in gender-typed appearance from 4 to 5 years, $r(176) = -.19$, $p = .011$. There were no

Table 2
Correlations Among Change Scores of Gender Rigidity Measures

Measure	1	2	3	4
Change between consecutive ages				
1. Appearance	—	.05	−.01	−.19*
2. Dress-up play	.12	—	−.06	−.12
3. Toy play	.00	−.06	—	.16*
4. Sex segregation	−.01	.17*	−.10	—

Note. $N = 175$. Correlations below the diagonal refer to changes in variables from ages 3 to 4 years. Correlations above the diagonal refer to changes in variables from age 4 to 5 years. See Halim (2012) for further details.
$^{*}p < .05$.

significant differences by ethnicity. However, in examining correlations by gender we found that for boys only, increases in gender-typed appearance were positively associated with increases in gender-typed play at the same time interval, $r_{3\text{–}4}(93) = .31$, $p = .002$; $r_{4\text{–}5}(93) = .20$, $p = .055$. Surprisingly, for girls these relations trended in the negative direction, $r_{3\text{–}4}(83) = -.21$, $p = .055$; $r_{4\text{–}5}(93) = -.19$, $p = .087$, and differed from boys' correlations ($p < .001$, $p = .004$, respectively).

Stability of Gender-Typing Within Measures

Stability in children's gender-typing was tested in the structural equation model as above (Figure 2) with gender-typing scores at one age predicting gender-typing scores at the consecutive age. Notably, both gender-typed dress-up play and gender-typed play showed considerable stability across consecutive ages (dress-up: $\beta_{\text{age3–4}} = .49$, $p < .001$, $\beta_{\text{age4–5}} = .40$, $p < .001$; play: $\beta_{\text{age3–4}} = .38$, $p < .001$, $\beta_{\text{age4–5}} = .53$, $p < .001$). Gender-typed appearance ($\beta = .22$, $p = .010$) and sex segregation ($\beta = .24$, $p = .012$) were only significantly related from ages 4 to 5 years. The estimates from ages 3 to 4 years were smaller than those from ages 4 to 5 years and were only a trend for appearance ($\beta = .14$, $p = .083$), but not significant for sex segregation ($\beta = .10$, $p = .237$). Multiple group analyses indicated no significant improvement in overall model fit when examining separate models by gender or ethnicity. However, examination of the data revealed that the stability of gender-typed appearance was evident for girls, even from ages 3 to 4 years, ($\beta_{\text{age3–4}} = .31$, $p = .001$, $\beta_{\text{age4–5}} = .32$, $p = .002$) but not for boys ($\beta_{\text{age3–4}} = .02$, $p = .894$, $\beta_{\text{age4–5}} = .13$, $p = .364$).

Discussion

Cognitive theories of gender development posit that as young children learn about gender categories and that they belong to one of these categories, they will be highly motivated to conform to gender stereotypes (Martin et al., 2002). As children actively construct and consolidate information about gender, their strict adherence may increase in rigidity (Ruble, 1994). Do gender-typed behaviors of young children follow this predicted pattern of increasing rigidity? The current results lend support to this conceptualization of gender development and indicate its robustness across children from different ethnic backgrounds. Each of the four gender-typed behaviors followed a course of change from ages 3 to 5 years, with nearly all behaviors moving toward greater rigidity. Specifically, children's dress-up play, gender-typed play, and sex segregation increased in rigidity from ages 3 to 4 years. Children's observed gender-typed appearance was already high at age 3, but did not increase in level. Findings point to the transition from ages 3 to 4 years as an especially important time in children's gender development for the active consolidation of gender information. This consolidation manifests in multiple aspects—from gendered activities to the peers with whom children play. The period from ages 4 to 5 years showed more divergence. Sex segregation trended toward increasing further, corroborating past, largely cross-sectional, research on middle-class White children (Ruble et al., 2006). In contrast, dress-up play and gender-typed play remained unchanged from ages 4 to 5 years, although cross-gender-typed play decreased. Gender-typed appearance decreased in rigidity, suggesting growing flexibility.

In light of the behavioral rigidity we observed, an interesting juxtaposition is that during the same developmental period children show flexibility in some gender-related judgments. For example, research has shown that a majority of preschoolers viewed exclusion from gender-stereotyped activities based on gender (e.g., girls excluding a boy from playing dolls with them because he is a boy) as unfair and wrong (Killen, Pisacane, Lee-Kim, & Ardila-Rey, 2001; Theimer, Killen, & Stangor, 2001). Furthermore, the judgment that exclusion based on gender is wrong becomes more pronounced, complex, and more based on moral reasoning rather than on social convention as children develop through elementary school (Killen, 2007). In addition, around age 5 or 6 years, children's gender-stereotype beliefs start to grow in their flexibility (Trautner et al., 2005). Thus, at times, children's gender-typed behaviors might misalign with their gender-related beliefs, perhaps due to feeling pressure from peers to conform to gender norms in their behavior. This is an interesting area for future investigation and highlights the importance of examining in future research whether the developmental shifts we observed are linked to children's growing understanding of gender, as cognitive theories would predict, as it is possible that other perspectives (e.g., socialization, biological) could also explain rigidity in gender-typed behaviors.

Children's physical appearance may be the first to move toward flexibility because it may peak in rigidity earlier as well, as physical appearance is one of the most salient traits to children in early childhood (Ruble & Dweck, 1995). Indeed, research on the "pink, frilly dress" phenomenon and girls' extremely intense interest in playing dress-up suggests that this appearance rigidity could start as early as age 2 (DeLoache, Simcock, & Macari, 2007; Halim et al., 2013). Gender-related cognitions, especially gender stereotypes, tend to increase in rigidity then grow more flexible (e.g., Trautner et al., 2005), and interestingly, in this same study, children who showed earlier rigidity showed earlier flexibility. Findings here suggest that the trajectory of gender-typed appearance is similar to that of gender-related cognitions.

Patterns in the timing and shape of trajectories for the four behaviors were robust across gender. That no gender differences in the timing or shape of any trajectory for any of the four behaviors were seen is somewhat surprising as some research on White, middle-class children would suggest that girls sometimes precede boys in gender development (La Freniere, Strayer, & Gauthier, 1984; Signorella et al., 1993; Zosuls et al., 2009). Perhaps in Latino and African American cultures, masculinity (i.e., *machismo*) is strongly emphasized so that gender development for boys occurs as early as it does for girls (e.g., DeSouza et al., 2004; Rowan, Pernell, & Akers, 1996).

In terms of ethnic variation, Dominican and African American children were similar in their developmental trajectories. Mexican children differed the most. For some measures (gender-typed appearance, dress-up play), Mexican children's trajectories suggest that their peaks in rigidity and subsequent flexibility in gender-typed behavior occurred earlier. The concordance in the trajectories of African American and Dominican children may be due to possessing similar demographics (e.g., mother's education, father cohabitation, urban background). Alternatively, more cross-cultural influence may occur between Dominican and African American groups because some Dominicans are perceived and/or identify as Black and also because of the influence of African American hip hop culture on Dominican youth (Duany, 2008; Torres-Saillant & Hernandez, 1998). The uniqueness of Mexican children's trajectories may have been due to less acculturation. Although both Dominican and Mexican families were largely first generation, Dominicans resided in the United States for more years than Mexicans, and a small proportion of them (15%) were second generation.

Group Similarities and Differences in Levels of Gender-Typing

A strength of this study was our inclusion of multiple measures of gender-typed behavior, as stressed by the matrix of gender-typing (Ruble et al., 2006). This strength became evident when we examined group differences in levels of gender-typing. If we had relied on only one measure, we would have erroneously concluded that one group was more gender-typed than the other, when, in fact, each group manifested gender rigidity in different behaviors.

Gender similarities and differences. Notably, we found differences in the ways that boys and girls expressed their gender. Girls engaged in dress-up play more than did boys. And although we could not formally test differences in gender-typed appearance, girls appeared to find a larger number of ways to dress in gender-stereotypical ways compared to boys. Thus in two of the four measures, girls' manifested their gender rigidity in self-adornment rather than in what they play with or with whom they play. Prior research has suggested that boys sometimes show stronger gender-typed behavior compared to girls, but these studies focused largely on play (Huston, 1983; Perry, White, & Perry, 1984; Turner, Gervais, & Hinde, 1993). Our results suggest that girls' adherence to maintaining a feminine appearance and in pretending to be princesses should also be examined as important gendered behaviors. Future research should examine the broader societal implications of this early emphasis on physical appearance for girls. If girls tie their self-worth to one's physical appearance this could potentially put girls at risk for psychological distress later in life (Crocker & Wolfe, 2001; Nolen-Hoeksema, 2001).

Consistent with past research, boys and girls did not differ in levels of sex segregation (Fabes, Martin, & Hanish, 2003; Howes, 1988). They also did not differ in levels of gender-typed play (including cross-gender-typed play), which may be surprising. Studies have shown that White, middle-class grade school boys compared to girls tend to reject cross-gender-typed behaviors and preferences more (Pickering & Repacholi, 2001; Raskin & Israel, 1981). The lack of a gender difference in our sample could possibly be due to the young age or to the different ethnic background and class of our sample. This is an area for future research.

Cultural similarities and differences. Another strength of this study was the diversity of the sample. We explored whether Latino children would have higher levels of gender-typed behaviors compared to African American children based on the literature suggesting egalitarian gender roles in African American families (e.g., Hill, 1971). Results largely indicated no differences in levels. Indeed, it was remarkable that African American and Latino children often looked the same. Both African American and Latina girls exhibited similar levels of dress-up play, gender-typed play, and sex segregation. And boys showed no ethnic group differences. These results indicate that rigidity in gendered behaviors is robust and generalizable to children of multiple ethnic backgrounds.

One exception was that Dominican girls dressed in more feminine ways than did Mexican girls. Qualitative interviews with Caribbean American communities have revealed persistent traditionalism in gender roles concerning household responsibilities and liberties (Lopez, 2002). These values could translate into the encouragement toward Dominican girls to maintain a highly feminine appearance. Of course, in interpreting differences, it is important to recognize that although groups in our sample were similar on a number of demographic variables and representative of the ethnic groups in New York City, a limitation of our study was that not all demographic variables were identical. Future research should attempt to equalize other demographic variables to determine the cause of these ethnic group differences.

Why we found no ethnic differences in boys is a question for future research. Perhaps both Latino and African American cultures value masculinity in boys to such a high degree (DeSouza et al., 2004; Rowan et al., 1996) that it would mask any cultural variation. Indeed, the boys often wore baggy clothing with contrasting red and black and graffiti designs suggesting an early "toughness" that may symbolize masculinity as young as age 3 years in an urban setting.

Coherence of Gender-Typing

Whether or not different dimensions of gender-typing show coherence within a single individual has long been debated (Huston, 1983). We found no significant correlations among the different behaviors within each age even though at the group level, three of the four behaviors increased in rigidity from ages 3 to 4 years. Our findings support studies suggesting that gender-typing dimensions are distinct, at least when assessed within age. When we examined change across successive years, a different picture emerged. Increases in sex segregation were positively correlated with increases in dress-up play and gender-typed play at certain time intervals consistent with past research on the bidirectional link between gender-typed play and sex segregation (Martin & Fabes, 2001). It was surprising that changes in gender-typed appearance showed no connections to changes in dress-up play and further surprising that changes in gender-typed appearance showed no positive connections to any other behavior for girls. In fact, increases in gender-typed appearance were related to decreases in gender-typed play for girls and decreases in sex segregation for both boys and girls from ages 4 to 5 years. Perhaps these negative or null relations are due to the disparate trajectories of each behavior with gender-typed appearance showing an earlier period of rigidity. Alternatively perhaps peers are rejecting children who look highly gender-stereotypical at later ages. This is a question for future research.

The different conclusions drawn from within-time versus across-time associations between different gender-typed behaviors may shed light on the conundrum of gender-typing coherence (Martin & Ruble, 2009). On one hand, the within-time data suggest that at one point in time, children may express their gender in a variety of ways. A child who dresses in an extremely feminine way does not necessarily play with only dolls and kitchen sets. The lack of associations may also indicate that different behaviors emerge at different times in response to different antecedents (e.g., cognitive development or social pressures). On the other hand, the across-time data suggest that gender-typing shows some unity with some gender-typed behaviors changing in parallel to others for certain groups and at certain age intervals. Overall then, gender-typed behaviors seem to exhibit specificity in one sense (within-time), but more cohesion in another (across-time). These data stress the importance of looking at associations between variables across time in future research.

Stability in Gender-Typing

Did children show stability in specific gender-typed behaviors across time? Or is gender-typing primarily a group-level rather than an individual-level phenomenon (Maccoby, 2002)? We found stability in dress-up play and gender-typed play across successive ages for the whole sample. Girls also showed stability in their gender-typed appearance. In addition, children's sex segregation was stable between the ages of 4 and 5 years. *R*s ranged from .18 to .47. These associations further support our claim that children may adopt individualized ways of displaying gender through play, appearance, and dress-up play that continue over time. Moreover, these stability coefficients were robust and generalizable across different ethnic groups.

The lack of stability for sex segregation from ages 3 to 4 years is consistent with some past research (Maccoby & Jacklin, 1974), but not with other research (Martin & Fabes, 2001; Serbin et al., 1993). Perhaps this is not surprising given that our research design was most consistent with Maccoby and Jacklin's (1987) study that examined children over a 2-year period from ages 4½ to 6½ years. The lack of sex segregation at age 3 years may partly explain this lack of stability—the mean percent of same-gender friends was not significantly different from 50%. This suggests that sex segregation in low-income African American and Latino populations emerges later than what is found in White, middle-class samples (around 27 months for girls, 36 months for boys; Ruble et al., 2006). On another note, the lack of stability in sex segregation from ages 3 to 4 years, and the lower level of stability from ages 4 to 5 years, compared to the other three behaviors, hints at the possibility that sex segregation may reflect different underlying processes. Perhaps sex segregation is less of a reflection of the internally motivated, child-directed consolidation of gender, and more of a reflection of socialization and of children's experiences in playing with the other gender, such as girls' dislike for boys' tendencies toward rough-and-tumble play (Pellegrini & Smith, 1998).

The lack of stability in appearance for boys is intriguing as well and opens up directions for future research. Perhaps appearances are not as central to boys in expressing their gender compared to other behaviors such as play (Miller, Lurye, Zosuls, & Ruble, 2009). It is also possible that girls get to choose their clothing at earlier ages than boys do. Thus, starting from age 3 years, girls' appearances may already reflect their preferences for feminine clothing, whereas boys' appearances may not reflect their own preferences until later ages.

These findings argue for continued research on individual-level differences in gender-typing, not just group-level ones. They suggest that even though gendered behaviors do not necessarily align in a constellation of gender-typing within a certain age, when just one behavior is examined, that behavior shows stability within an individual child over time. Future research should examine whether these individual differences during ages 3–5 years have downstream influences on children's identity, behavior, preferences, and values later in life. For example, it would be interesting to see whether girls who are low on gender-typing tend to become tomboys in middle childhood or whether girls who are high on gender-typing tend to disengage with mathematics and science in adolescence. These future directions would further speak to possible antecedents of gender-typing and allow us to further examine stability within the context of normative increasing rigidity followed by flexibility later in development.

Limitations

It is important to note that mothers provided data for several of the behaviors we examined, with the exception of gender-typed appearance. We stress the importance of continued research on these behaviors using more observational techniques to replicate our results. In addition, although the developmental trajectories we observed support the hypotheses of cognitive theories of gender development, it is important to recognize other perspectives on gender development, such as biological and socialization perspectives (Ruble et al., 2006). Parents' own adherence to gender norms may influence the importance they place on children's gender conformity, and families make decisions that influence what their children wear, what they play with, and with whom they play. Nevertheless, some studies have shown that parent influence on children's appearances is minimal across ages 3–6 years (Halim et al., 2013), and children are likely to have considerable independence from parents in play and peer choices in preschool contexts. However, exploring other influences is an important area of future research.

In addition, another limitation of our study was our inability to examine acculturation's influence on gender-typed behavior (Chuang & Tamis-LeMonda, 2009). In our sample, ethnicity was confounded with immigration status as all African

American children were third generation or later, and all Latino children were first or second generation. Even within the Latino group, the vast majority of children had foreign-born parents, so we were unable to compare first- versus second-generation children. It would be interesting for future research to examine whether acculturation plays a role in the level and timing of gender-typing. For example, it is possible that since American culture values equality, including gender equality (Tamis-Lemonda & McFadden, 2009), later generation children would show lower levels of gender-typing and later peaks in gender-typing. At the same time it is important to note that all of the children in our sample were part of a large urban context, and indeed, even part of a larger U.S. context. The levels and patterns of gender-typed behavior we observed may reflect this larger context as well.

Conclusion

A child can express gender in multiple behaviors—through what they wear, through the costumes they don, through the objects they play and do not play with, and through the peers that they befriend. On a group level, most of these behaviors generally showed similar trajectories from ages 3 to 4 years, with children increasingly acting in gender-stereotypical ways, consistent with predictions based on cognitive theories of development. From ages 4 to 5 years, the behaviors diverged. The timing and levels of behavior sometimes varied by ethnic group, highlighting the importance of studying gender development in multiple, diverse groups. Despite group-level parallels, at the individual level, children who were highly gender-typed in one way (e.g., wearing a lot of pink dresses) were not necessarily gender-typed in another way (e.g., playing frequently with dolls). Nevertheless, there was some coherence across pairs of gender-typed behaviors in change across time, as well as moderate stability in individual gendered behaviors from year to year. In short, our findings contribute to the ongoing conversation about the developmental trajectories, multidimensional nature, and stability of individual differences of gender-typing in an ethnically diverse sample through a longitudinal design. The examination of young children's gender-typed behaviors provided a snapshot into a dynamic, exciting period of gender development, involving both stability and change that could have implications for gender over the life course.

References

Albert, A. A., & Porter, J. R. (1988). Children's gender-role stereotypes: A sociological investigation of psychological models. *Sociological Forum, 3,* 184–210. doi:10.1007/BF01115290

Bailey, B. A., & Nihlen, A. S. (1990). Effect of experience with nontraditional workers on psychological and social dimensions of occupational sex-role stereotyping by elementary school children. *Psychological Reports, 66,* 1273–1282. doi:10.2466/PR0.66.4.1273-1282

Chuang, S. S., & Tamis-LeMonda, C. S. (2009). Gender roles in immigrant families: Parenting views, practices, and child development. *Sex Roles, 60,* 451–455. doi:10.1007/s11199-009-9601-0

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review, 3,* 593–625. doi:10.1037/0033-295X.108.3.593

DeLoache, J. S., Simcock, G., & Macari, S. (2007). Planes, trains, automobiles–and tea sets: Extremely intense interests in very young children. *Developmental Psychology, 43,* 1579–1586. doi:10.1037/0012-1649.43.6.1579

DeSouza, E. R., Baldwin, J., Koller, S. H., & Narvaz, M. (2004). Latin American perspective on the study of gender. In M. A. Paludi (Ed.), *The psychology of gender* (pp. 41–68). Westport, CT: Praeger.

Duany, J. (2008). *Quisqueya on the Hudson: The transnational identity of Dominicans in Washington Heights.* New York: CUNY Dominican Studies Institute.

Fabes, R. A., Martin, C. L., & Hanish, L. D. (2003). Young children's play qualities in same-, other-, and mixed-sex peer groups. *Child Development, 74,* 921–932. doi:10.1111/1467-8624.00586

Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental trajectories of sex-typed behavior in boys and girls: A longitudinal general population study of children aged 2.5-8 years. *Child Development, 79,* 1583–1593. doi:10.1111/j.1467-8624.2008.01207.x

Halim, M. L. (2012). *Gender-typed behavior in ethnically diverse young children: Developmental trajectories, cognitive underpinnings, and consequences for gender identity and attitudes* (Unpublished doctoral dissertation). New York University, New York.

Halim, M. L., Ruble, D. N., Lurye, L., Greulich, F., Zosuls, K., & Tamis-LeMonda, C. S. (2013). *The case of the pink frilly dress and the avoidance of all things "girly": Girls' and boys' appearance rigidity and cognitive theories of gender development.* Manuscript submitted for publication.

Hill, R. B. (1971). *The strengths of Black families.* New York: Emerson Hall.

Howes, C. (1988). Relations between child care and schooling. *Developmental Psychology, 24,* 53–57. doi:10.1037/0012-1649.24.1.53

Huston, A. C. (1983). Sex-typing. In E. M. Hetherington (Ed.), *Handbook of child psychology: Socialization, personality, and social development* (Vol. 4, pp. 387–467). New York: Wiley.

Jarret, R., Roy, K., & Burton, L. (2002). Fathers in the "hood": Insights from qualitative research on low-income African-American men. In C. S. Tamis-LeMonda & N. Cabrera (Eds.), *Handbook of father involvement: Multidisciplinary perspectives* (pp. 211–248). Mahwah, NJ: Erlbaum.

Killen, M. (2007). Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science*, *16*, 32–36.

Killen, M., Pisacane, K., Lee-Kim, J., & Ardila-Rey, A. (2001). Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Developmental Psychology*, *37*, 587–596. doi:10.1037/0012-1649.37.5.587

La Freniere, P., Strayer, F. F., & Gauthier, R. (1984). The emergence of same-sex affiliative preferences among preschool peers: A developmental/ethological perspective. *Child Development*, *55*, 1958–1965. doi:2155/10.2307/1129942

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation. *Monographs of the Society for Research in Child Development*, *67*(2, Serial No. 269). doi:10.1111/1540-5834.00189

Lopez, N. (2002). *Hopeful girls, troubled boys: Race and gender disparity in urban education*. New York: Routledge.

Maccoby, E. E. (2002). Gender and group process. *Current Directions in Psychological Science*, *11*, 54–58. doi:10.1111/1467-8721.00167

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Maccoby, E. E., & Jacklin, C. N. (1987). Sex segregation in childhood. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 239–288). New York: Academic Press.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, *37*, 431–446. doi:10.1037/0012-1649.37.3.431

Martin, C. L., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: Children's beliefs about playing with girls versus boys and their relations to sex segregated play. *Journal of Social and Personal Relationships*, *16*, 751–771. doi:10.1177/0265407599166005

Martin, C. L., & Halverson, C. F., Jr. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, *52*, 1119–1134. doi:2256/10.2307/1129498

Martin, C. L., & Ruble, D. N. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, *13*, 67–70. doi:10.1111/j.0963-7214.2004.00276.x

Martin, C. L., & Ruble, D. N. (2009). Patterns of gender development. *Annual Review of Psychology*, *61*, 353–381. doi:10.1146/annurev.psych.093008.100511

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*, 903–933. doi:10.1037/0033-2909.128.6.903

McArdle, J. J. (2009). Latent variable modeling of differences and changes in longitudinal data. *Annual Review of Psychology*, *60*, 577–605. doi:10.1146/annurev.psych.60.110707.163612

Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles*, *60*, 870–881. doi:10.1007/s11199-009-9584-x

Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science*, *10*, 173–176. doi:10.1111/1467-8721.00142

Parsons, A., & Howe, N. (2006). Superhero toys and boys' physically active and imaginative play. *Journal of Research in Childhood Education*, *20*, 287–300. doi:10.1080/02568540609594568

Pellegrini, A. D., & Smith, P. K. (1998). Physical active play: The nature and function of a neglected aspect of play. *Child Development*, *69*, 577–598. doi:10.1111/j.1467-8624.1998.tb06226.x

Perry, D. G., White, A. J., & Perry, L. C. (1984). Does early sex typing result from children's attempts to match their behavior to sex role stereotypes? *Child Development*, *55*, 2114–2121. doi:10.2307/1129784

Pickering, S., & Repacholi, B. (2001). Modifying children's gender-typed musical instrument preferences: The effects of gender and age. *Sex Roles*, *45*, 623–643. doi:10.1023/A:1014863609014

Raskin, P. A., & Israel, A. C. (1981). Sex-role imitation in children: Effects of sex of child, sex of model, and sex-role appropriateness of modeled behavior. *Sex Roles*, *7*, 1067–1077. doi:10.1007/BF00287584

Rowan, G. T., Pernell, E. J., & Akers, T. A. (1996). Gender role socialization in African American men: A conceptual framework. *Journal of African American Men*, *1*, 3–22. doi:10.1007/BF02733916

Ruble, D. N. (1994). A phase model of transitions: Cognitive and motivational consequences. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 163–214). New York: Academic Press.

Ruble, D. N., & Dweck, C. S. (1995). Self-perceptions, person conceptions, and their development. In N. Eisenberg (Ed.), *Social development: Review of personality and social psychology* (Vol. 15, pp. 109–139). Thousand Oaks, CA: Sage.

Ruble, D. N., Lurye, L. E., & Zosuls, K. M. (2007). Pink frilly dresses (PFD) and early gender identity. *Princeton Report on Knowledge (P-ROK)*, 2(2). Retrieved January 2, 2013, from http://www.princeton.edu/prok/issues/2-2/pink_frilly.xml

Ruble, D. N., & Martin, C. L. (1998). Gender development. In W. Damon (Ed.), *Handbook of child psychology* (Vol. 3, 5th ed., pp. 933–1016). New York: Wiley.

Ruble, D. N., Martin, C., & Berenbaum, S. (2006). Gender development. In W. Damon (Ed.), *Handbook of child psychology* (Vol. 3, 6th ed., pp. 858–932). Hoboken, NJ: Wiley.

Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development, 78,* 1121–1136. doi:10.1111/j.1467-8624.2007.01056.x

Serbin, L. A., Powlishta, K. K., & Gulko, J. (1993). The development of sex-typing in middle childhood. *Monographs of the Society for Research in Child Development, 58*(2, Serial No. 232).

Servin, A., Bohlin, G., & Berlin, L. (1999). Sex differences in 1-, 3-, and 5-year-olds' toy-choice in a structured play-session. *Scandinavian Journal of Psychology, 40,* 43–48. doi:10.1111/1467-9450.00096

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review, 13,* 147–183. doi:10.1006/drev.1993.1007

Stanback, M. H. (1985). Language and Black woman's place: Evidence from the Black middle class. In P. A. Treichler, C. Kramarae, & B. Stafford (Eds.), *For alma mater: Theory and practice in feminist scholarship* (pp. 177–193). Urbana: University of Illinois Press.

Tamis-Lemonda, C. S., & McFadden, K. E. (2009). The United States of America. In M. H. Bornstein (Ed.), *Handbook of cultural developmental science* (pp. 299–322). New York: Psychology Press.

Theimer, C. E., Killen, M., & Stangor, C. (2001). Young children's evaluation of exclusion in gender-stereotypic peer contexts. *Developmental Psychology, 37,* 18–27. doi:10.1037/0012-1649.37.1.18

Torres-Saillant, S., & Hernandez, R. (1998). *The Dominican Americans*. Westport, CT: Greenwood.

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in children: Developmental or differential? *Infant and Child Development, 14,* 365–381. doi:10.1002/icd.399

Turner, P. J., Gervais, J., & Hinde, R. A. (1993). Gender-typing in young children: Preferences, behavior and cultural differences. *British Journal of Developmental Psychology, 11,* 323–342. doi:10.1111/j.2044-835X.1993.tb00607.x

U.S. Census Bureau. (2011). *Statistical abstract of the United States, population*. Retrieved April 13, 2011, from http://www.census.gov/compendia/statab/2011/tables/11s0005.pdf

Zosuls, K. M. (2008). *Gender development in African American, Dominican, and Mexican immigrant toddlers: Self socialization in three cultural contexts* (Unpublished doctoral dissertation). New York University, New York.

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for sex-typed play. *Developmental Psychology, 45,* 688–701. doi:10.1037/a0014053

Child Development, July/August 2013, Volume 84, Number 4, Pages 1285–1295

Forces and Motion: How Young Children Understand Causal Events

Tilbe Göksun
University of Pennsylvania

Nathan R. George and
Kathy Hirsh-Pasek
Temple University

Roberta M. Golinkoff
University of Delaware

How do children evaluate complex causal events? This study investigates preschoolers' representation of *force dynamics* in causal scenes, asking whether (a) children understand how single and dual forces impact an object's movement and (b) this understanding varies across cause types (Cause, Enable, Prevent). Three-and-a half- to 5.5-year-olds (n = 60) played a board game in which they were asked to predict the endpoint of a ball being acted upon by one or two forces. Children mostly understood the interactions of forces underlying each type of cause; only 5.5-year-olds could integrate two contradictory forces. Children perceive force interactions underlying causal events, but some concepts might not be fully understood until later in childhood. This study provides a new way of thinking about causal relations.

Consider the everyday events of an ordinary playground: children throwing Frisbees, blocking soccer balls, and helping each other push their friends on the swing. Psychologists tend to describe these complex scenes with simple causal mechanisms (Michotte, 1963) or chains of simple causes (e.g., Baillargeon, 1994; Cohen, Rundell, Spellman, & Cashon, 1999; Leslie, 1982, 1984; Oakes & Cohen, 1990; Rakison & Krogh, 2012; Saxe, Tenenbaum, & Carey, 2005; Saxe, Tzelnic, & Carey, 2007). In this article, we move beyond this simple Michottean causality and examine how children process complex causal scenes in light of force dynamics theory. Defined as the interaction between entities in space resulting from multiple forces (Talmy, 1988; Wolff, 2003, 2007), *force dynamics* encompasses not only simple cause–effect relations but also ways in which two or more forces can affect the trajectory of an entity in an event. Drawing on Wolff's (2003, 2007) force dynamics theory of causal language, we investigate the development of force dynamics representations in children and reconsider the psychologist's definition of causality given the complexity of the causal relations that constitute everyday events.

Force Dynamics

Until recently, force dynamics has been "a neglected field of study" (Talmy, 1988) in semantics, as well as in spatial understanding, event perception, and language development. Event perception research has traditionally emphasized the perception of causality from a Michottean perspective (Michotte, 1963), relying on direct causal events in which one object (i.e., agent) contacts a stationary object (i.e., patient), causing the patient to immediately move along the agent's trajectory while the agent ceases movement. The spatiotemporal contiguity between the two movements creates the perception of causality. Evidence for the primacy of spatiotemporal features in causal perception has yielded promising results with both adults (e.g., Choi & Scholl, 2006; Fugelsang, Roser, Corballis, Gazzaniga, & Dunbar,

---

This work was supported by NICHD Grant 5R01HD050199 and by NSF Grant BCS-0642529 to the second and third authors. We thank everyone at the Temple University Infant Lab and in the Spatial Intelligence and Learning Center (particularly the RISC group at Temple University) for their invaluable contributions to this project. We also thank Nora Newcombe for helpful suggestions on many phases of this project. Special thanks to Kelly Fisher, Sarah Roseberry, Andrea Frick, Wendy Shallcross, Yannos Misitzis, Russell Richie, Katrina Ferrara, Carolyn Winslow, and Rhea Miles for their help in stimulus design, data collection, and coding. We are grateful to the children and parents who participated in the study.

Correspondence concerning this article should be addressed to Tilbe Göksun, Department of Neurology, University of Pennsylvania, 3400 Spruce Street, 3 Gates, Philadelphia, PA 19104. Electronic mail may be sent to tilbe@mail.med.penn.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0013
DOI: 10.1111/cdev.12035

2005; Schlottmann, Ray, Mitchell, & Demetriou, 2006) and infants (e.g., Carey, 2009; Cohen & Amsel, 1998; Leslie & Keeble, 1987; Muentener & Carey, 2010; Newman, Choi, Wynn, & Scholl, 2008; Oakes, 1994; Rakison & Krogh, 2012; Saxe et al., 2005; Saxe et al., 2007; Shultz, 1982). Many have used such findings to support a mechanist account of causal perception, in which the transmission of some physical quantity (e.g., energy, momentum) is considered necessary and sufficient for the perception of causal interactions (see Shultz, 1982).

While such conclusions might reflect a valid part of causal knowledge, they are limited with respect to the full array of causal interactions. Other factors relevant to the interpretations of complex causal events have not been explained. For example, a vital component of a causal event, the spatial array of forces (e.g., the angle of an object's movement after collision), has recently been shown to be influential to causal perception in adults (Straube & Chatterjee, 2010; Wolff, 2007). Moreover, previous research has focused primarily on one-force causal interactions, yet the environment often contains multiple kinetic forces working in concert. Extending the analysis of causality into these more complicated scenarios may reveal a greater range of development in these representations.

Building on work by Talmy (1985, 1988) and Jackendoff (1990), a recent theory addresses the limitations of previous mechanist accounts of causality (Wolff, 2003, 2007). The *force dynamics model* makes improvements on previous research by taking into account the multiple forces that underlie the full array of causal events. Force dynamics expands our view of cause, distinguishing among the *types* of cause: *Cause* (one force that moves an object), *Enable* (a secondary force that promotes motion in the intended direction), and *Prevent* (a secondary force that hinders motion in the intended direction; Wolff, 2003). (We choose "enable" rather than "help" to be in accord with the force dynamics literature; e.g., Sloman, Barbey, & Hotaling, 2009; Wolff, 2003, 2007.) For example, when a boat is moving toward a port, a secondary force, "the wind," might *prevent* the boat from doing so by changing its direction. In contrast, if the boat were hit by a ship moving in the intended direction, the force of the ship would *enable* the boat to reach port. Wolff and colleagues found that adults correctly map causal verbs onto events through the identification of these three event classes (Cause, Enable, and Prevent), supporting the theory that causality has different category-level representations (Wolff, 2003, 2007; Wolff & Song, 2003; Wolff, Song, & Driscoll, 2002). It appears that these causal categories are codified in a number of languages. For example, Wolff, Klettke, Ventura, and Song (2005) showed that English, German, Russian, Spanish, and Arabic all use verbs to express Cause, Enable, and Prevent relations. The authors concluded that these expressions of causal relations are not unique to English. Here, we borrow Wolff's terminology of the force dynamics of causal language and adjust it to assess preschool children's perception of forces and their interactions in complex causal scenes.

Understanding an Object's Trajectory and Integrating Forces in Childhood

Infants in their 1st year of life expect motion events to have causes (e.g., Harding & Golinkoff, 1979) and have some knowledge of an object's trajectory in such events, even when a portion of the object's path is occluded (Johnson et al., 2003; Kotovsky & Baillargeon, 2000; Muentener & Carey, 2010; Spelke, Philips, & Woodward, 1995). Yet preschool children often have misconceptions about the trajectories of an object's motion when causal forces are combined. For example, 4.5- to 5.5-year-old children predicted a straight path rather than a parabolic trajectory when asked to combine momentum and gravity to judge the path of a ball rolling off a table or dropping from a moving train (Kaiser, McCloskey, & Proffitt, 1986; Kaiser, Proffitt, & McCloskey, 1985; but see Hood, 1995, 1998). Moreover, research has shown that even older children have trouble integrating forces to estimate the movement of an object and consequently the result of an event. For example, Pauen (1996) presented elementary schoolchildren (ages between 6 and 10 years) a task in which two forces acted on a target object. The target object was fixed at the center of a circular platform, surrounded by a barrier with an opening. This barrier could be rotated around the platform. Each force was represented by different number of weights resting on small plates connected to the target object by separate cords. In each trial, the experimenter manipulated the number of weights on the plates (magnitude) and the angle between the cords (direction). Children were asked to rotate the barrier to position the opening such that the target would slide through it. Results showed that children younger than 10 years of age seemed to ignore one aspect of the vectors—either magnitude or direction. One intriguing exception is preschool children's reasoning about the path of a ball exiting a curved tube. Five-year-olds correctly solve this common physics problem, predicting a straight path for the ball; however, following entry into school, children begin attributing a curved path to the ball, suggesting that such early reasoning is not the robust, sophisticated reasoning seen in adults (Kaiser et al., 1986). Such fallacies about the relations between force and motion persist even into high school (e.g., McDermott & Redish, 1999; Roth, 1990).

To this point, research on force dynamics in childhood has mostly been restricted to knowledge of intuitive physics and nonverbal causal reasoning (for a review, see Wilkening & Cacchione, 2011). However, there is value in also examining how force dynamics is manifested both in event representations and in the corresponding semantic categories of language that both go beyond simplistic pictures of causality. Recent evidence suggests that before producing motion verbs such as *cross* or *jump*, children detect the components of motion events (e.g., the trajectory of the movement, the figure of the action, the manner of the figure). That is, they acquire the necessary conceptual basis before they lexicalize different types of events in languages (Göksun, Hirsh-Pasek, & Golinkoff, 2010b). Similar conclusions can be drawn about the development of simple causal language.

As Wolff et al. (2005) suggested, many languages represent these complex events in their causal language system. Developmental research has shown that children start using causal verbs such as *break* and *cut* in the 2nd year of life, but it is not until age 4 years that they reliably use causal verbs and other causal connectives to describe causal relations (Bowerman, 1974; Clark, 2003; Göksun, Hirsh-Pasek, & Golinkoff, 2010a). Interestingly, 4- and 5-year-olds rely on gestures to supplement their speech before they form complete causal sentences. For example, these children represent the instrument of a causal relation only in gesture without mentioning it in speech (Göksun et al., 2010a). Thus, the productive use of causal language lags behind the understanding of simple causes.

According to one view, learning to use different types of causal verbs (Cause, Enable, and Prevent) to talk about causal scenes that integrate various forces requires that children understand complex causal events. Wolff's (2007) study with adults suggests some characteristics that these representations may contain. His research has shown that the representations of force dynamics underlying language encode both the direction and the magnitude of forces working in concert. Alternatively, exposure to verbs that encode Cause, Enable, and Prevent, may encourage children to attend to complex, multiforce events. In other words, although simple causal language may be based on early concepts, learning causal language might help children to move from this basic understanding of causal relations to force dynamic representations (for arguments on a different domain-learning number words and counting, see Lipton & Spelke, 2005). Thus, the examination of children's understanding of multiple force events sets the stage for determining whether these representations are precursors to force dynamics verbs or whether these semantic categories enhance children's representations of these events.

The Current Study

Force dynamics defines different classes of causal events and verbs according to different patterns of forces. To make these distinctions among causal events, children must be able to combine and integrate multiple forces. The current state of the field is not clear on whether young children can combine and integrate multiple forces before they productively talk about complex causal relations (Clark, 2003). This study examines how preschoolers evaluate multiple forces and process events in terms of force dynamics. We ask three questions: First, do children between 3.5 and 5.5 years of age understand the impact of one and two forces on an object's movement and its endpoint? Based on the findings of Pauen (1996), we hypothesize that younger children might be unable to integrate two forces to predict the movement of an object. Second, are children's predictions equally accurate among the different types of cause (Cause, Enable, Prevent)? We predict that children might represent Cause and Enable relations earlier than Prevent relations, because Prevent involves two *contradictory* forces. Finally, what types of errors do children make when trying to integrate forces? We analyze children's errors for each categorical representation to see whether children perseverate on a single force, despite the presence of other forces.

Method

Participants

The final sample consisted of 60 monolingual English-speaking children, balanced for gender and evenly separated into three age groups: 3.5-year-olds (*M* = 44.64 months, *SD* = 2.67, range = 40.02–48.08 months), 4.5-year-olds (*M* = 54.95 months, *SD* = 3.94, range = 49.06–60.21 months), and 5.5-year-olds (*M* = 64.86 months, *SD* = 2.36, range =

61.17–71.0 months). Children were predominantly White and of middle-class families from a Northeastern city in the United States. Based on prior research on children's understandings of other physical problems (e.g., Kim & Spelke, 1999), and children's effective use of causal language (e.g., Göksun et al., 2010a), these age groups were chosen to represent the complete developmental trajectory for force dynamics. Additional data from five 3.5-year-olds and one 4.5-year-old were excluded from analysis due to failure to respond during the task (six). In addition, data from one child from each of the older groups were discarded due to experimenter error (two). Finally, nine adults ($M = 28.77$ years, $SD = 4.91$) were recruited to assess adults' performance on this novel task.

Materials

Children's understanding of causal relations was examined through an interactive board game. The game board (180 cm × 180 cm) was made of a green felt mat and divided into a 6 × 6 grid of 30 cm × 30 cm squares using masking tape (see Figure 1). A 1.5-in. diameter curved plastic tube (16 in. tall × 5 in. deep) was inserted into a model house so that it entered into the roof and exited out from the front. In all conditions, the first force was provided by the motion of a small foam ball exiting the tube. In Enable and Prevent relations, the hairdryer's "wind" was used as a secondary force. There were also four colored felt squares (30 cm × 30 cm). The child was asked to choose the one they liked for marking the final location of the ball. The board was laid on the ground in a quiet room where each child was tested individually. A camera recorded the session for later coding.

Procedure

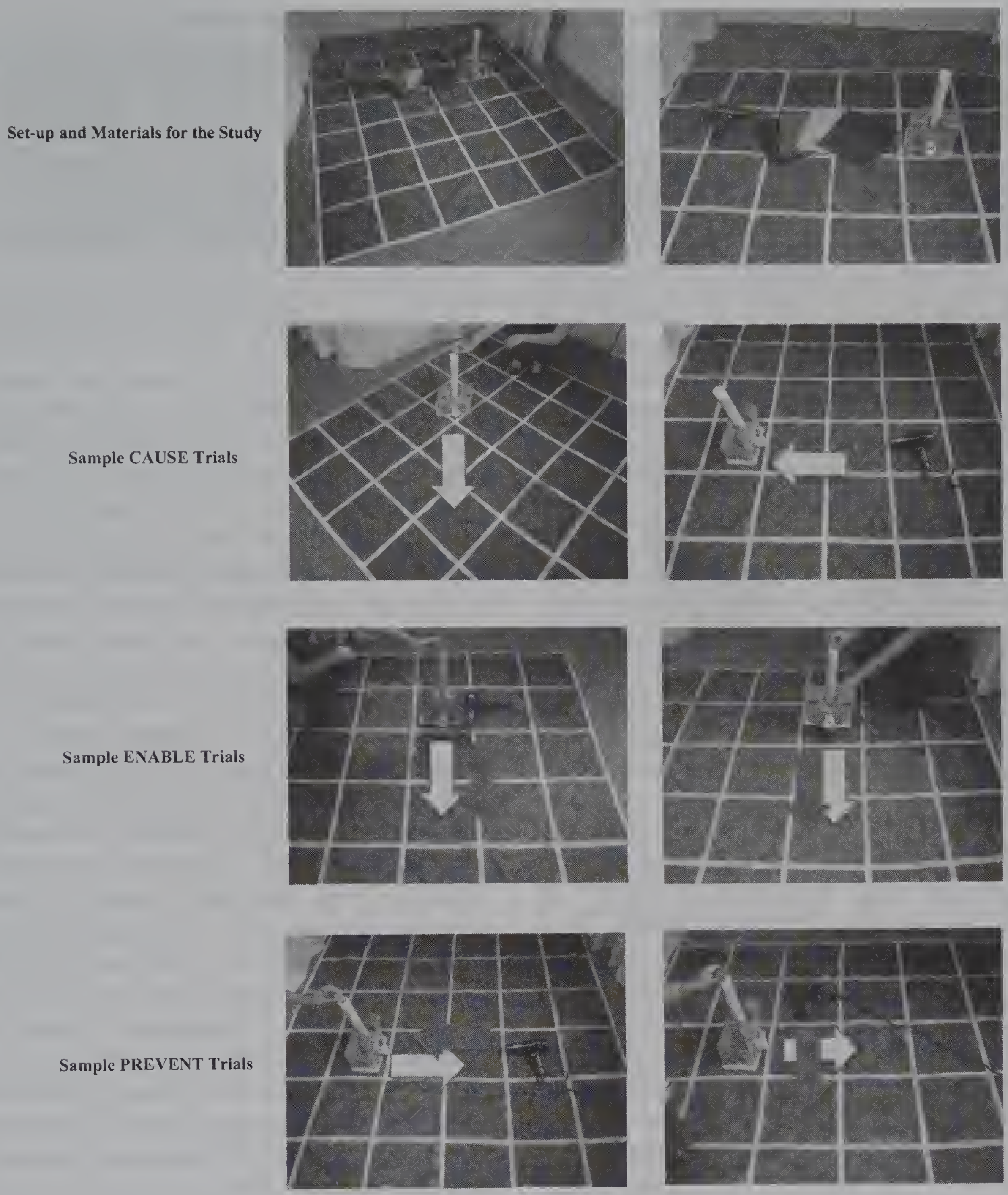
The experiment consisted of two test phases (one- and two-force trials). Each test phase started with a practice trial. The order of the phases (one- and two-force trials) remained constant among children.

One-force trials: Practice trial. First, the experimenter presented the task materials and described the game to the child (see Figure 1, top row). Next, the experimenter asked the child to pick her favorite colored square and explained the child's role in the game: "The goal of the game is to guess where the ball will stop. You are the player and will guess where the ball goes and stops on these squares." Then, the house was put on one of the board's grid squares, with the end of the tube in line with the edge of the next square. The experimenter drew the child's attention to the tube saying, "Wow, do you see the long chimney and how it goes through the house and comes out the side?" Finally, the experimenter took the ball and held it over the top of the tube, asking the child where she thought the ball would exit and how far it would travel. After the child put the colored square on the felt mat, the experimenter said, "Watch what happens when I drop the ball from here (the top of the tube). Did you see how it goes? See how it came out from here (showing the curved part of the tube—the door of the house)?" Next, the experimenter asked the child to drop the ball from the tube so that the child was familiar with the procedure and the weight of the ball.

After being familiarized with the task materials, the experimenter placed the house on a different square and asked the same question as before, now requiring the child to put the colored square (the same size as one of the grid squares) on the grid square where she thought the ball would land. The child was permitted to place the colored square anywhere, including locations off the board, although such responses were not encouraged from the outset. The child was asked to step aside and told to watch where the ball stopped. The experimenter dropped the ball from the chimney and the ball stopped on the board. If the child was incorrect, the experimenter would encourage the child to try again and would repeat the action one more time. The practice trial ensured that the child was familiar with the materials and attuned to the movement of the ball with one force. If a child failed in the second practice trial, the experimenter explained the procedure one more time and ensured that the child understood the task by asking the child to point at the end state of the ball.

One-force trials: Test trials. Phase 1 of the test contained three simple Cause trials, in which the experimenter was the only cause of the action (*one-force* trials). The house was placed in different positions and turned in either horizontal or diagonal directions as shown in Figure 1. By manipulating the orientation of the house, we examined whether children's integration of forces was dependent on any specific spatial orientation. The child either received two horizontal trials and one diagonal trial or two diagonal trials and one horizontal trial in a counterbalanced order. For each location, the child made a prediction using the colored square and the experimenter dropped the ball through the tube.

Two-force trials: Practice trial. This phase started by introducing the child to the workings of the hairdryer that served as the second force (i.e., the wind) on the ball. The child felt the "wind" by



Set-up and Materials for the Study

Sample CAUSE Trials

Sample ENABLE Trials

Sample PREVENT Trials

Figure 1. The board game setup with the model house, hairdryer, and materials (top row) and sample Cause, Enable, and Prevent trials (one- and two-force trials). The yellow arrows represent the direction of the ball exiting the model house (initial force) and the red arrows represent the direction of the hairdryer (secondary force). In prevent trials, the left picture is the "opposing" Prevent and the right picture is the "perpendicular" Prevent. See the online version of this article for Figure 1 in color.

turning the hairdryer on and placing his or her hand close to the opening. Then, the experimenter put the ball in front of the hairdryer on the mat and turned the hairdryer on to show the child that the wind from the hairdryer moved the ball on the mat. After this introduction and ensuring that the child was still engaged in the game, the house and the hairdryer were positioned on the board for a practice trial. The hairdryer was lined up flush to the middle of the left edge of the square in front (15 cm) of the fort, with the wind blowing across the front of the house. The experimenter turned on the hairdryer and then rested his hand on top of the tube. Then, the child was asked where the ball would land and to put his or her colored square on the grid. After the child placed his or her colored square, the experimenter dropped the ball through the tube. The child was again asked to step aside and told to watch where the ball stopped. If a child was incorrect, the experimenter would encourage the child to try again and repeat the action one more time. The experimenter always ensured that the child followed where the ball landed by asking the child to point at the end state of the ball. In this second practice trial for two forces, the experimenter confirmed that the child saw the movement of the ball with the second force.

Two-force trials: Test trials. One Cause trial with the hairdryer, two Enable trials, and two Prevent trials were presented in a counterbalanced order using the same script as above. In the new Cause trial, the hairdryer was located two squares in front of the house (60 cm) and pointed back toward the house (in this case the hairdryer is the Cause). Both the edge of the tube and the end of the hairdryer were lined up centered and flush with the edge of a grid square. The ball was put in the center of the square directly in front of the hairdryer (15 cm) and the child was asked to predict the endpoint of the ball when only the hairdryer acted upon it.

In Enable trials, the house was placed flush with the edge of a grid square and the hairdryer was either placed next to the house (15 cm from center of tube to center of hairdryer) or under the house (the center of the tube 10 cm above the center of the hairdryer). The ball exiting the house was propelled along its initial line of motion by the hairdryer in either case. Each child received one diagonal and one horizontal trial.

In Prevent trials, the ball's motion upon exiting the house was precluded by the hairdryer, located either opposite the house ("opposing Prevent") or perpendicular to the house ("perpendicular Prevent"). In the opposing Prevent, the hairdryer was two squares in front of the house (60 cm) facing back toward the house. In perpendicular prevent trials, the hairdryer was lined up flush to the middle of the left edge of the square in front (15 cm) of the fort, with the wind blowing across the front of the tube. These relations were both classified as Prevent because they prevented the ball from continuing on its original of motion. In the Enable and Prevent trials, both the experimenter and the hairdryer were forces acting upon the ball (*two-force* trials; see Figure 1 for sample trials).

Coding and Reliability

The accuracy of children's responses was coded for each trial. To count as correct, children's predictions for both the ball's direction and endpoint were considered. Figure 2 presents the correct and incorrect responses in each trial. For example, a horizontal Cause trial was counted as correct if the child guessed that the ball would stop on the board right in front of the house. That is, a correct response included the prediction that the ball would stop on the first or second grid square in front of the house, as was seen by the child in the warm-up trial. A correct answer in the horizontal Enable trial included the prediction that the ball's endpoint was at least the third grid square from the house, indicating recognition that the additional force should propel the ball farther than in Cause trials. For the perpendicular Prevent trial in which the hairdryer was located perpendicular to the house, the child's guess should have been 45° to the right of the ball's original path. For the opposing Prevent trial, the correct answer should be the grid square directly in front of the house. Although correct responses varied by trial in the number of squares that are considered correct, we believed any resulting bias is minimal. Specifically, the nature of the scheme was categorical, not exact, and so whether the category of "farther than the cause trial" or "shorter than the cause trial" contained one, two, or three possible locations should be of minimal importance. Finally, we coded children's exact predicted square in relation to the house.

Children's errors were classified into different types based on the trial. For Cause trials, an error was scored as "no straight path" if the child does not take the ball's straight motion into account (e.g., predicting the ball would turn sharply left or right) or other (e.g., predicting an area outside the board). For Enable trials, the errors were divided into two types: "no straight path" as before or "no wind's force" if the child does not take the wind's force into account (i.e., not predicting an endpoint beyond the Cause trial result). For Prevent trials, the errors could be related to either the straight path or the wind's force, or both.

To obtain intercoder reliability, 20% of children were coded by a second person who was blind to the task. Agreement between coders was 96% ($k = .94$, $n = 96$).

Results

Preliminary Analyses

First, we analyzed adults' correct predictions to one- and two-force trials. Paired-sampled t tests showed that adults' responses ($M_{\text{one-force}} = 100$ and $M_{\text{two-force}} = 88.90$) were at ceiling for one-force trials and almost at ceiling for two-force trials, yielding no significant difference between trial types, $t(8) = 1.84$, $p > .05$. Further analysis indicated no reliable differences between Enable and Prevent trials, $p > .05$ ($M_{\text{Enable}} = 88.90$ and $M_{\text{Prevent}} = 88.90$). Thus, these results suggest that adults could judge the direction and endpoint of both one- and two-force trials, successfully integrating two forces in this novel task.

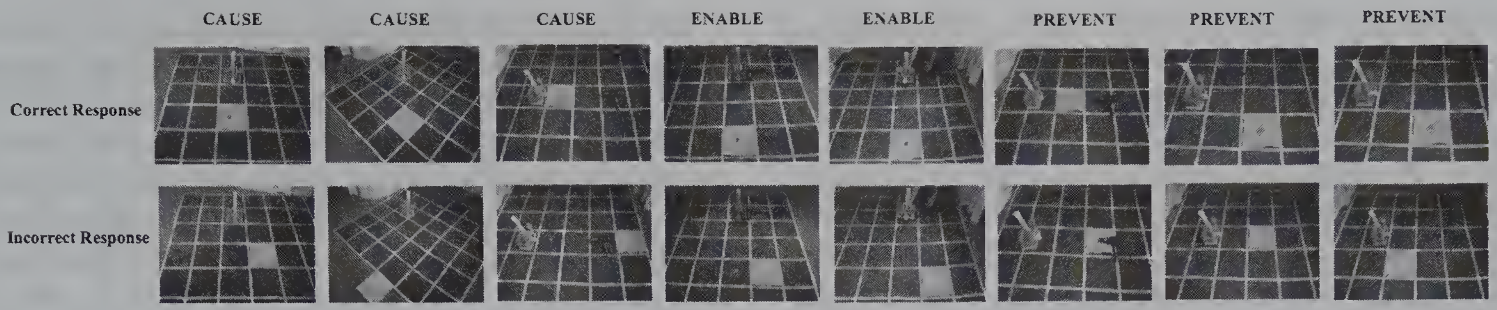


Figure 2. Sample correct and incorrect responses from a set of trials in the task. The yellow square in each trial demonstrates a child's choice. See the online version of this article for Figure 2 in color.

Second, a repeated measure analysis of variance (ANOVA) with age (3.5-, 4.5-, and 5.5-year-olds) and gender as between-subject variables, and causal type (Cause, Enable, Prevent) as a within-subject variable indicated no significant main effect of gender or any interactions with gender. Thus, gender was not considered in further analyses.

Predictions in One- and Two-Force Trials

Children's mean percentage of correct responses to one- and two-force trials was analyzed using a repeated measure ANOVA with age (3.5-, 4.5-, and 5.5-year-olds) as a between-subject variable, and trial type (one- vs. two-force) as a within-subject variable. Results yielded significant main effects of age and trial type as well as an Age × Trial Type interaction, $F(1, 57) = 44.09$, $p < .01$, $\eta_p^2 = .34$; $F(2, 57) = 4.69$, $p < .01$, $\eta_p^2 = .14$; and $F(2, 57) = 3.92$, $p < .05$, $\eta_p^2 = .12$, respectively. As shown in Figure 3, post hoc analyses indicated that 5.5-year-olds gave reliably more correct responses than 3.5- and 4.5-year-olds only in two-force trials (Scheffé, *ps* < .03). Further pair-sampled *t* tests indicated that 5.5-year-olds were not significantly different from adults in their predictions of one-force or two-force trials, $t(27) = 1.75$, $p > .05$ and $t(27) = 2.63$, $p > .05$, respectively.

Next we analyzed whether the orientation of the model house in relation to the board (horizontal vs. diagonal) affected children's responses for one- versus two-force trials. For both types of trials, children in all age groups responded significantly better in horizontal directions, $t(59) = 3.35$, $p < .05$, and $t(59) = 2.56$, $p < .05$, respectively (one-force: $M_{\text{horizontal}} = 93\%$ and $M_{\text{diagonal}} = 73\%$ and two-force: $M_{\text{horizontal}} = 68\%$ and $M_{\text{diagonal}} = 53\%$).

Predictions in Cause, Enable, Prevent

Children's responses to different cause types (Cause, Enable, Prevent) showed main effects of age and cause type, $F(2, 57) = 6.28$, $p < .01$, $\eta_p^2 = .32$, and $F(2, 114) = 28.57$, $p < .01$, $\eta_p^2 = .21$. No interaction between age and cause type was found (see Figure 4). Overall, 5.5-year-olds were better than both 3.5- and 4.5-year-olds in predicting the correct direction and endpoint of the ball in all types of trials (Scheffé, *ps* < .05). A further analysis indicated that 5.5-year-olds were not significantly different from adults, except in the Prevent trials, $t(27) = 2.10$, $p < .05$. Children in all age groups were better at predicting the results of Cause compared to Enable and Prevent trials, $t(59) = 2.053$, $p < .05$, and $t(59) = 7.63$, $p < .05$, for Cause–Enable and Cause–Prevent comparisons, respectively. However, results of Enable trials seemed easier to

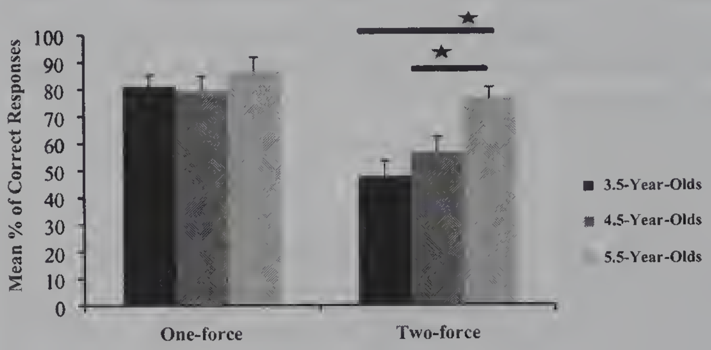


Figure 3. Mean percentages of children's correct responses to one-force and two-force trials. The error bars present the standard errors of the means.
*$p < .05$.

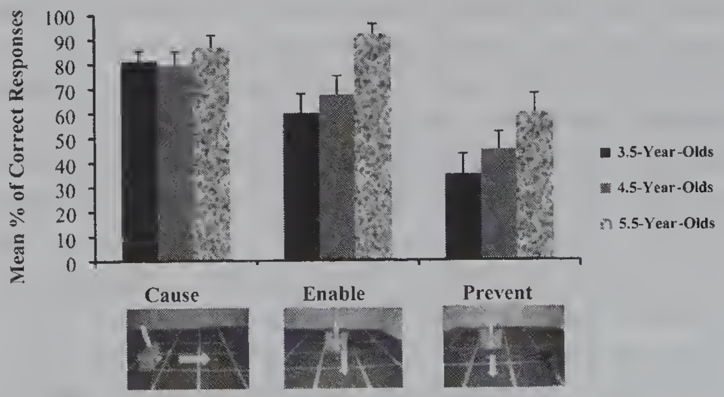


Figure 4. Mean percentages of children's correct responses to Cause, Enable, and Prevent relations.

predict than Prevent trials for both horizontal and diagonal orientations of the model house, $t(59) = 3.09$, $p < .05$, and $t(59) = 3.43$, $p < .05$ (Enable: $M_{\text{horizontal}} = 80\%$ and $M_{\text{diagonal}} = 67\%$ and Prevent: $M_{\text{horizontal}} = 55\%$ and $M_{\text{diagonal}} = 38\%$). Moreover, children in all age groups found the perpendicular Prevent trial harder than the opposing Prevent trial, $t(59) = 3.59$, $p = .01$ ($M_{\text{perpendicular}} = 34\%$ and $M_{\text{opposing}} = 62\%$). Finally, we analyzed children's exact prediction location for Cause and Enable events. Results indicated that regardless of the age group, children added the wind's force in Enable trials and predicted a location further away from the house than in Cause trials, $F(1, 52) = 79.71$, $p < .0001$, $\eta_p^2 = .61$.

Errors in Prediction

When children erred in the Cause trials they did not consider that the ball should move in a straight path after being dropped into the tube. For Enable relations, younger children made similar prediction errors not taking the straight path into account. Nevertheless, the very few errors 5.5-year-olds made for Enable relations were the result of discounting the second force (the wind) on the ball's motion. That is, although these children predicted the path correctly, they placed the colored square on the first square in front of the house instead of farther out. Finally, children in all age groups rarely accounted for the second force (i.e., the wind) and perseverated on the person's force on the ball's motion for Prevent. Adults made very few errors in the Prevent and Enable relations and all were the result of only taking the second force into account.

Discussion

This study is among the first to go beyond children's understanding of simple cause and ask how they might integrate the intertwined effects of multiple causal forces. Following Wolff's theory of force dynamics, we asked preschool-aged children to predict the direction and endpoint of an object in three types of causal relations (Cause, Enable, and Prevent) that constitute the basis for the force dynamics model. Overall, results suggested that to some extent children represent force dynamics in causal events. We found that: (a) children were better at judging the direction and endpoint of one-force trials, with only 5.5-year-olds able to integrate two forces; (b) children at all ages were poorest with Prevent trials; and (c) children disregarded the secondary force (i.e., wind) when judging the endpoint of two-force interactions.

To some extent, children's knowledge of causal events extends beyond simple collisions among physical objects to a range of possible agents working in concert. The number and direction of forces in causal relations are crucial in predicting the goal of the action. The current study shows that while children are good at reasoning about one-force causal events, they are selectively attentive to secondary forces, incorporating them only if both forces move in the same direction. This characteristic of children's reasoning reflects their first, simplified rule for integrating forces. Prior research suggests that when asked to judge a rectangle's area, 3- and 4-year-olds use an additive integration rule (adding the height and width of a rectangle) rather than multiplying them (Cuneo, 1982; see also Schlottmann & Anderson, 1994; Wilkening, 1981; but see Siegler's (1976) work on balance-beam problems). In addition to these studies, our study suggests that contrary to Piaget's notion of centration (Piaget, 1955), preschoolers can focus on two dimensions of a task using the additive integration rule. Indeed, only 5.5-year-olds developed beyond this additive strategy predicting the endpoint of the ball for some, but not all, of the two-force relations. They are not yet at adult levels when forces oppose one another.

The results of the Prevent trials shed light on the factors that influence force integration in childhood. In particular, the perpendicular Prevent relations were very difficult for preschool children, likely because the secondary force is at a 90° angle from the initial force. In contrast, when the Prevent event is similar to the Enable one (i.e., when two forces interact by working in the same, singular dimension), children make fewer erroneous predictions. It seems likely, therefore, that the number of dimensions involved in multiple force interactions is an important factor in prediction above and beyond the concordance of forces. This may represent the most difficult hurdle to overcome in the current task.

Previous research demonstrates that children have similar misconceptions in other domains (for a review, see Wilkening & Cacchione, 2011). For example, 4-year-olds have difficulty combining gravity and inertia when predicting where a ball would land after it rolled from a slightly downward slanted ramp. These children predict that the ball would descend in a straight path whereas 6-year-olds correctly predict the ball's parabolic path (Kim & Spelke, 1999). Hood (1995) presented children with a vertical invisible displacement task, in which they were asked to find a ball dropped down one of the chimneys connected to three hiding boxes. Before 3 years of age, children did not take the trajectory of the tubes into account and searched for the ball in the box that was directly connected to the chimney where the ball was dropped. In tasks that require the calculation of different dimensions of vectors (weight and angle), even 9- and 10-year-olds have difficulty integrating dimensions (Pauen, 1996). Taken together, our findings provide converging evidence that before age 5, children might perseverate on a single dimension when predicting the direction of an object's movement. Only after this age do they start considering the second dimension in certain contexts. Thus, children's conceptions about different variables in physical reasoning (gravity and inertia, integration of dimensions of vectors, and force dynamics) might not emerge until the later elementary school years.

These findings might also reflect an ongoing developmental trajectory in other cognitive domains such as memory and spatial reasoning. Our task not only requires children to integrate vectors and reason about causality but also to store different dimensions in memory and to consider the spatial orientation of the model house (horizontal vs. diagonal). Results from the orientation of the house show that children have difficulty with the diagonal task, possibly related to their emerging spatial reasoning and memory capacity. However, spatial reasoning per se may not be the problem. Children as young as 2 years of age remember the locations of several hidden objects sequentially (Sluzenski, Newcombe, & Satlow, 2004) and found the location of an object when hidden at one location in a circular enclosure (Balcomb, Newcombe, & Ferrara, 2011). Future studies should tease apart the role of these cognitive abilities in evaluating force dynamics among preschoolers.

Caution is required in interpreting the current findings. One question of interest is whether children at this age have enough experience with wind as conveyed by a hair dryer. Informal discussions with parents suggested that the children had experiences with hairdryers and that this was not an issue, but a more systematic investigation is warranted. In addition, in the beginning of the two-force trials, we assured that children felt the wind by turning the hairdryer on and placing their hands close to the opening. Even with these assurances, however, we do not know whether another type of a second force (e.g., another ball) would lead to similar predictions.

Second, for this study, we chose to employ a board game task involving predictions, a rather stringent test of force dynamics. The possibility remains that preschoolers have a basic knowledge of even complex interactions that was not captured by the strict measures employed. Future experiments that present choices for the endpoint of the ball (e.g., will the ball stop on this square or this square?), and do not require children to generate a response, may further enhance our understanding of the state of preschoolers' knowledge of force dynamics.

Finally, our results hint at the relation between causal concepts and children's learning of complex causal language. Prior research shows that children do not productively use a full cadre of causal verbs to describe causal actions before ages 4–5 (Bowerman, 1974; Clark, 2003; Göksun et al., 2010a). To use different force dynamics verbs (e.g., *Cause:* cause, make, set; *Enable:* help, allow, let; *Prevent:* block, stop, keep), children might first need to differentiate and conceptualize various causal events represented by force dynamics. Nevertheless, the use of various force dynamics verbs might also direct children's attention to relevant characteristics of multiple force interactions. That is, language might be a flashlight on force dynamics and might help children form categories of force dynamics in much the same way that number words helps children organize their understanding of counting (e.g., Lipton & Spelke, 2005). Either way, we expect a tight link between the causal events that children reason about and the events that they readily describe using force dynamics verbs. Future studies should examine how and at what age children encode these differences in their linguistic descriptions of events, and the nature of the relation between linguistic and nonlinguistic force dynamics representations.

In conclusion, our findings provide evidence about how preschoolers use force and motion to evaluate the causal relations they witness. Complex causal relations involving the interaction of forces are difficult to interpret before 5 years of age. Finally, this research opens an avenue to examine the relation between children's understanding of force dynamics and their use of a variety of causal verbs in language.

References

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, *3*, 133–140. doi:10.1111/j.2044-835X.1994.tb00616.x

Balcomb, F., Newcombe, N. S., & Ferrara, K. (2011). Finding where and saying where: Developmental relationships between place learning and language in the second year. *Journal of Cognition and Development*, *12*, 315–331. doi:10.1080/15248372.2010.544692

Bowerman, M. (1974). Learning the structure of causative verbs: A study in the relationship of cognitive, semantic and syntactic development. *Papers and Reports on Child Language Development*, *8*, 142–178.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception*, *35*, 385–399. doi:10.1068/p5462

Clark, E. V. (2003). *First language acquisition*. New York: Cambridge University Press.

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of causality. *Infant Behavior and Development*, *21*, 713–731. doi:10.1016/S0163-6383(98)90040-6

Cohen, L. B., Rundell, L. J., Spellman, B. A., & Cashon, C. H. (1999). Infants' perception of causal chains. *Psychological Science*, *10*, 412–418. doi:10.1111/1467-9280.00178

Cuneo, D. O. (1982). Children's judgments of numerical quantity: A new view of early quantification. *Cognitive Psychology*, *14*, 13-41018-1029. doi:10.1016/0010-0285(82)90003-2

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, *24*, 41–47. doi:10.1016/j.cogbrainres.2004.12.001

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010a). How do preschoolers express cause in gesture and speech? *Cognitive Development*, *25*, 56–68. doi:10.1016/j.cogdev.2009.11.001

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010b). Trading spaces: Carving up events for learning language. *Perspectives on Psychological Science*, *5*, 33–42. doi:10.1177/1745691609356783

Harding, C. G., & Golinkoff, R. M. (1979). The origins of intentional vocalizations in prelinguistic infants. *Child Development*, *50*, 33–40.

Hood, B. M. (1995). Gravity rules for 2–4 year-olds? *Cognitive Development*, *10*, 577–598. doi:10.1016/0885-2014(95)90027-6

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, *1*, 59–63. doi:10.1111/1467-7687.00013

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, *74*, 94–108. doi:10.1111/1467-8624.00523

Kaiser, M. K., McCloskey, M., & Proffitt, D. R. (1986). Development of intuitive theories of motion in the absence of external forces. *Developmental Psychology*, *22*, 67–71. doi:10.1037/0012-1649.22.1.67

Kaiser, M., Proffitt, D. R., & McCloskey, M. (1985). The development of beliefs about falling objects. *Perception & Psychophysics*, *38*, 533–539. doi:10.3758/BF03207062

Kim, I. K., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science*, *2*, 339–362. doi:10.1111/1467-7687.00080

Kotovsky, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5-month-old infants. *Developmental Science*, *3*, 344–359. doi:10.1111/1467-7687.00129

Leslie, A. M. (1982). The perception of causality in infants. *Perception*, *11*, 173–186. doi:10.1068/p110173

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, *13*, 287–305. doi:10.1068/p130287

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*, 265–288. doi:10.1016/S0010-0277(87)80006-9

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, *76*, 978–988. doi:10.1111/j.1467-8624.2005.00891.x

McDermott, L. C., & Redish, E. F. (1999). Resource letter PER-1: Physics education research. *American Journal of Physics*, *67*, 755–767.

Michotte, A. E. (1963). *The perception of causality*. New York: Basic Books.

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*, 63–86. doi:10.1016/j.cogpsych.2010.02.001

Newman, G. E., Choi, H., Wynn, K., & Scholl, B. J. (2008). The origins of causal perception: Evidence from postdictive processing in infancy. *Cognitive Psychology*, *57*, 262–291. doi:10.1016/j.cogpsych.2008.02.003

Oakes, L. (1994). Examining in infancy: Does it reflect active processing? *Developmental Psychology*, *30*, 869–879. doi:10.1037/0012-1649.30.6.869

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*, 193–207. doi:10.1016/0885-2014(90)90026-P

Pauen, S. (1996). Children's reasoning about the interaction of forces. *Child Development*, *67*, 2728–2742. doi:10.1111/j.1467-8624.1996.tb01885.x

Piaget, J. (1955). *The child's conception of the world*. London: Routledge & Kegan Paul.

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science*, *15*, 43–53. doi:10.1111/j.1467-7687.2011.01096.x

Roth, K. J. (1990). Developing meaningful conceptual understanding in science. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 139–176). Hillsdale, NJ: Erlbaum.

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*, 995–1001. doi:10.1111/j.1467-9280.2005.01649.x

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43*, 149–158. doi:10.1037/0012-1649.43.1.149

Schlottmann, A., & Anderson, N. A. (1994). Children's judgments of expected value. *Developmental Psychology, 30*, 56–66. doi:10.1037/0012-1649.30.1.56

Schlottmann, A., Ray, E. D., Mitchell, A., & Demetriou, N. (2006). Perceived physical and social causality in animated motions: Spontaneous reports and ratings. *Acta Psychologica, 123*, 112–143. doi:10.1016/j.actpsy.2006.05.006

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1, Serial No. 194).

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520. doi:10.1016/0010-0285(76)90016-5

Sloman, S., Barbey, A. K., & Hotaling, J. M. (2009). A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science, 33*, 21–50. doi:10.1111/j.1551-6709.2008.01002.x

Sluzenski, J., Newcombe, N., & Satlow, E. (2004). Knowing where things are in the second year of life: Implications for hippocampal development. *Journal of Cognitive Neuroscience, 16*, 1443–1451. doi:10.1162/0898929042304804

Spelke, E. S., Philips, A. T., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–77). Oxford, England: Oxford University Press.

Straube, B., & Chatterjee, A. (2010). Space and time in perceptual causality. *Frontiers in Human Neuroscience, 4*, 28.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 57–149). New York: Cambridge University Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49–100. doi:10.1016/0364-0213(88)90008-0

Wilkening, F. (1981). Integrating, velocity, time, and distance information: A developmental study. *Cognitive Psychology, 13*, 231–247. doi:10.1016/0010-0285(81)90009-8

Wilkening, F., & Cacchione, T. (2011). Children's intuitive physics. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 473–496). Chichester, UK: Wiley.

Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition, 88*, 1–48. doi:10.1016/S0010-0277(03)00004-0

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*, 82–111. doi:10.1037/0096-3445.136.1.82

Wolff, P., Klettke, B., Ventura, T., & Song, G. (2005). Categories of causation across cultures. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside of the lab: Festschrift in Honor of Douglas L. Medin* (pp. 29–48). Washington, DC: American Psychological Association.

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*, 276–332. doi:10.1016/S0010-0285(03)00036-7

Wolff, P., Song, G., & Driscoll, D. (2002). Models of causation and causal verbs. In M. Andronis, C. Ball, H. Elston, & S. Neuval (Eds.), *Papers from the 37th Meeting of the Chicago Linguistics Society, Main Session* (Vol. 1, pp. 607–622). Chicago: Chicago Linguistics Society.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1296–1307

Sociocultural Settings Influence the Emergence of Prelinguistic Deictic Gestures

Dorothe Salomo
Max-Planck-Research Group Communication before Language and Max-Planck-Institute for Psycholinguistics

Ulf Liszkowski
Max-Planck-Research Group Communication before Language, Max-Planck-Institute for Psycholinguistics, and University of Hamburg

Daily activities of forty-eight 8- to 15-month-olds and their interlocutors were observed to test for the presence and frequency of triadic joint actions and deictic gestures across three different cultures: Yucatec-Mayans (Mexico), Dutch (Netherlands), and Shanghai-Chinese (China). The amount of joint action and deictic gestures to which infants were exposed differed systematically across settings, allowing testing for the role of social–interactional input in the ontogeny of prelinguistic gestures. Infants gestured more and at an earlier age depending on the amount of joint action and gestures infants were exposed to, revealing early prelinguistic sociocultural differences. The study shows that the emergence of basic prelinguistic gestures is socially mediated, suggesting that others' actions structure the ontogeny of human communication from early on.

Western-based research has identified infants' prelinguistic interaction and gestural communication skills as one of the most important milestones in early social and language development. First, early forms of triadic, object-related joint actions ("joint engagement"; Bakeman & Adamson, 1984) help infants identify verbal reference to objects (Tomasello & Todd, 1983). These triadic bouts of joint action are developmentally related to infants' word learning (Tomasello & Farrar, 1986). Second, infants communicate with triadic, object-related gestures ("deictic gestures") in complex ways, including the comprehension and production of communicative, referential, and social intentions (for a recent overview, see Liszkowski, 2010). These deictic gestures are developmentally related to infants' acquisition of first words (Carpenter, Nagell, & Tomasello, 1998; Harris, Barlow-Brown, & Chasin, 1995) and to the emergence of syntactic word–word combinations (Iverson & Goldin-Meadow, 2005).

However, still little is known about the very emergence of these prelinguistic interactions and gestures, especially from a sociocultural perspective. More cognitively oriented accounts emphasize that before sufficient cognitive skills are in place, social–interactional input is not doing much work (although it will thereafter; Tomasello, 2008. Social-constructivist accounts instead suggest that many behaviors are acquired socially through joint engagement and social scaffolding in the first place (Carpendale & Lewis, 2004; Vygotsky, 1978). The central question is whether infants' gestures emerge independently of social interaction and only enable or constitute interaction and later verbal exchange, or whether infants' gestures emerge through social–interactional experiences in the first place. Recent studies have established correlations between caregiver and infant gesture use (Liszkowski, Brown, Callaghan, Takada, & De Vos, 2012; Liszkowski & Tomasello, 2011; Rowe & Goldin-Meadow, 2009), but the directionality has remained unclear. It is currently unknown whether caregivers gesture in response to infants' otherwise independently developing gestures or, alternatively, caregivers' interactions and gestures lead infants to gesture. In this respect, cultural comparisons are informative for development (Whiting, 1980). Humans do not

---

We thank Lin Dongxian, Yi Salomo, Laura Shneidman, and Margret van Beuningen for their help with recruiting families. Thanks to Mireille Hassemer, Marloes van der Goot, Cornelia Seitz, and Merel Fokkema for data coding. We owe special thanks to the families in Mexico, the Netherlands, and China who offered us insights into their daily lives and allowed us to observe them at home.

Correspondence concerning this article should be addressed to Ulf Liszkowski, Max Planck Research Group Communication before Language, Max-Planck-Institute for Psycholinguistics, MPI for Psycholinguistics wundlaan 1, 6525 XD Nijmegen, Gelderland, The Netherlands. Electronic mail may be sent to ulf.liszkowski@mpi.nl.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2012/8404-0014
DOI: 10.1111/cdev.12026

develop outside cultural contexts, and socialization is a pervasive feature of human development. If there were differences across different sociocultural settings in how interlocutors engage young infants in joint action and deictic gestural communication, this would provide the grounds for a natural experiment to test the effects of interactional 'input' on the early ontogeny of prelinguistic gestures.

Several studies indeed suggest large variation in the sociocultural settings within which children develop. Seminal comparative work measured 3- to 10-year-old children's socialization and interaction across different cultures, showing how specific socioculturally defined settings shape children's social development (Whiting & Whiting, 1975). Further studies have shown that language socialization differs vastly across cultures, with some cultures apparently lacking infant-directed speech or believing that talking to preverbal infants is inappropriate (Pye, 1986; Schieffelin & Ochs, 1986; for an overview, see Lieven & Stoll, 2009). There are also substantial cultural differences in parental beliefs and practices on how to raise children, ranging from early promotion of development (see Weisner, 2009) to its rather passive acceptance (e.g., Lancy, 2010). Further, differences in socialization goals regarding the independence and interdependence of children's relations to their families (Kagitcibasi, 2005) have qualitative effects on how parents interact with their infants already in the first few months of life, with more proximal, tactile face-to-face interactions corresponding to interdependent socialization goals, and more distal, visual–verbal interactions corresponding to independent socialization goals (e.g., Keller, 2007; Keller, Otto, Lamm, Yovsi, & Kärtner, 2008).

Most of these studies have documented differences in the sociocultural settings within which infants develop (for an overview, see Hewlett, 1996), how these differences in socialization practices are rooted in cultural belief systems (Harkness & Super, 1996), or how they affect development once children fully participate in the respective socialization practices and are competent linguistic interactants (e.g., Feiring & Lewis, 1981). However, it has remained unknown whether sociocultural differences already impact the early, prelinguistic emergence of object-directed deictic gestures. Given that young infants' social environments vary substantially across cultural settings, systematic comparisons could thus provide the quasi-experimental conditions to investigate whether different social–interactional inputs affect the very emergence of prelinguistic gestures.

In the current study, we tested the hypothesis that social–interactional input as part of enculturation processes influences the emergence of prelinguistic gestures. We aimed at showing that socialization does not only impact children's development once they engage with language in the social practices of their cultures, but that enculturation impacts already the very emergence of prelinguistic gestures. To test this hypothesis, we first established systematic differences in input measures across three different sociocultural settings by quantifying the amount of deictic gestures directed at infants, and the amount of time others engaged infants in object-directed joint actions. As a test, we then compared infants' use of deictic gestures across these cultural groups. We measured infants' production of deictic gestures that emerge some time between 8 and 15 months of age (Carpenter et al., 1998), specifically index-finger pointing. Index-finger pointing is a milestone in the early bidirectional understanding of referential communication (as indicated by Western-based research; Liszkowski & Tomasello, 2011), and it is used communicatively in infants across various sociocultural settings (Liszkowski et al., 2012). However, it is still unknown whether index-finger pointing emerges as a consequence of social–interactional input, or whether it emerges independently and only results in interactional responses (Carpendale & Carpendale, 2010). We chose three specific cultural settings to test our hypothesis: Yucatec-speaking Mayans in Yucatán Mexico, Mandarin-speaking Chinese in Shanghai, and Dutch-speaking Netherlanders in Nijmegen.

We chose Netherlanders as a reference sample that is comparable to most of the findings in the developmental literature based on Western European cultural settings. Socialization practices in the Netherlands are child centered, with caregivers providing attention for the child, but within limits of caregivers' interest and availability, and with the goal that children also learn to play by themselves (Harkness, Super, & van Tijen, 2000). The other two cultures were chosen because their socialization practices do not only differ from European cultures but also from one another. Other than in most Western cultural settings, Mayans' cultural belief is that young children develop and learn largely independently of caregivers' behavior toward them (Gaskins, 2000). Anthropological reports of Mayan societies indeed suggest that caregivers hardly engage with their infants verbally (Brown, 2000; Pye, 1992) or in joint object play (Gaskins, 1996). Thus, under the hypothesis that socialization influences prelinguistic interaction, we expected less infant gestures in the Mayan compared to the Dutch sample. In contrast, Shanghai Chinese infants are at the center of attention, especially since the implementation of the one-child policy (Goh & Kuczynski, 2010; Ho, 1989). Parents and grandparents are anxious to ensure that their only child receives the best possible care (Fong, 2007; Goh, 2006), and the societal system is generally gauged toward ever earlier education and promotion of development and achievement. It has been documented that Chinese parents, compared to their U.S. counterparts, gesture relatively more for their 4-year-olds, possibly reflecting their heightened interest in teaching and instructing them (Goldin-Meadow & Saltzman, 2000). Thus, here we expected a difference in the opposite direction, with more infant gestures than in the Dutch sample. Any other possible pattern of differences between the three groups, albeit difficult to interpret, would reject the proposed socialization hypothesis. The null hypothesis of finding no differences between groups would be amenable to several interpretations and, for example, support a more cognitively oriented account, suggesting that socialization processes are not effective before infants gesture and actively participate in social exchanges (e.g., Tomasello, 2008).

We made natural observations of infants' daily activities. Natural observations are ecologically more valid than observations in predefined contexts, like laboratory-based experiments or elicitation tasks, in that they are more sensitive to capturing natural group differences. We conducted online scan samplings on different times of a day to determine how much of their time infants were naturally engaged in joint actions by any interactant (e.g., parents, relatives, adults, siblings, peers). Scan sampling is an established paper-and-pencil method in ethology and cultural anthropology, which is minimally intrusive and involves registering online the occurrence of predefined categories of behavior in a given time frame. In addition, we made video recordings on different times of a day to analyze in detail the forms and frequencies of infants' and their interlocutors' gestural communication.

Under the assumption that socialization affects the emergence of prelinguistic communication, we expected the Yucatec Mayans and Shanghai Chinese on the opposite extremes, with the Dutch covering some middle ground. In a first set of analyses we tested whether our three samples indeed differed from one another in the input they provided to their infants. We tested whether (a) Yucatec Mayans engaged their infants in less triadic interactions and used less deictic gestures for their infants compared to the Dutch, and (b) Shanghai Chinese engaged their infants more often in triadic interaction and used more deictic gestures for their infants compared to the Dutch. In a second set of analyses we looked at infants' deictic gestures as an outcome measure. We tested whether the predicted cultural group differences would pertain to infants' gesture use and, in particular, to their index-finger pointing. Further, we tested for positive correlations between the social–interactional input measures and infants' gestures. Importantly, differences on the cultural group level would be informative for the directionality of such correlations, as there was no a priori basis for differences between the age-matched infants across the three cultural groups that would follow our predicted pattern.

We coded also other, more primary child-care activities, which we expected to be similar across cultural groups. Further, we coded for the proximity of caregivers as another factor that could possibly influence interaction patterns (e.g., Feiring & Lewis, 1981). We chose an age interval that ranged from a period before to after the emergence of joint attentional behaviors and deictic gestures (as identified by Western-based research) to investigate the possibility of different developmental time courses across the cultural settings.

Method

Participants

Sixteen Yucatec Mayan infants (7 girls; mean age = 11 months 29 days, range = 8.01–15.25), 16 Dutch infants (9 girls; mean age = 11 months 23 days, range = 8.04–15.03), and 16 Chinese infants (9 girls; mean age = 11 months 26 days, range = 8.09–15.15) and their families and temporary interlocutors participated in the study. Infants were matched for age across all three groups following the initial data collection in Mexico. In each setting, half of the sample was between 8 and 11 months of age (the younger group) and half were between 12 and 15 months of age (the older group).

Yucatec Mayan infants were recruited in two small neighboring villages near the town Chemax on the Yucatán Peninsula (Mexico) with the help of a native assistant and several local people. The observer (first author) lived for about 6 weeks in a family in one of the focal villages. All families with infants in the appropriate age range were invited to participate in the study, and most families agreed. Infants in the Netherlands were recruited in Nijmegen, a medium-sized Dutch city, by the same observer who had lived there for half a year. Parents had previously registered in a database to participate in infant studies and were contacted by phone. In China, infants were recruited in the metropolis Shanghai. The same observer spent about 5 weeks in Shanghai living with a local relative who assisted her in the recruitment process. Chinese families with infants in the appropriate age range were approached in various city quarters by two local Mandarin-speaking assistants and asked to participate. The families in all three cultures received a monetary gift for participation as well as a DVD with the video recordings of their session.

Sample Characteristics

The Yucatec Mayan families lived in traditional villages, remote from tourism centers and other cities, with no public transport. Mayan infants typically grow up in large families. All infants had between 6 and 11 siblings—apart from one infant who was a firstborn. Older siblings often took care of their younger brothers and sisters. Almost all the men in the two participating villages were subsistence farmers raising mainly corn for their own family's consumption while women's primary responsibility was to run the house and the garden. Children's daily activities were primarily structured by their parents' work, which had priority and should not be interrupted by children (see Gaskins, 1999). There was little attention on how infants spent their time, with almost no adult interference, as long as they were safe. Mayan parents believe that young children develop and learn largely independently of caregivers' behavior toward them. Parents define their caregiver role in terms of ensuring the safety and well-being of the infant and in keeping the infant content by responding to its needs and desires, for example by soothing through frequent breastfeeding, but not in terms of stimulating interaction, modeling, or direct teaching (Gaskins, 2006).

Dutch families were chosen with the constraint that infants spent maximally 1 day per week in a day care, to ensure that home data collection was representative and comparable. Families had one to four children ($M = 2.3$).

In most Chinese families ($N = 14$), both parents worked and the grandparents took care of the child, which is a common pattern (Kun, 2007). Grandparents usually lived in the same apartment, forming a three-generation household. Four families employed a nanny for additional support. Single children in today's China are the center of attention, whose needs and interest are to be met above all else. This child-centered orientation has also been depicted in the media as the "little-emperor-syndrome." Children enjoy an elevated position at home, with multiple caregivers revolving around them (Goh & Kuczynski, 2010).

Design and Procedure

Our study consisted of two independent parts: (a) scan sampling of daily activities and (b) video analyses of gestures. Each participating family was visited four times—twice for the scan sampling portion of the study, each lasting 2 hr (totaling 4 hr), and twice for the video recordings, each lasting 30 min (totaling 1 hr). The four visits were timewise as contiguous as possible depending on the schedule of each family. The four visits represented random shots of daily life with the one constraint that infants had to be awake. If this was not the case, the observer postponed the meeting. For the Mayans, all four visits took place within 7 days, for the Dutch within 12 days, and for the Chinese within 8 days.

All data were collected by the same person (first author) who was not a member of either of the cultural groups. In all three groups a local assistant who spoke the native language was present for the introduction at the first appointment. Families were informed that we were interested in how infants spent their everyday life in different parts of the world. They were not informed about any of our hypotheses or specific dependent measures. Terms like *social interaction*, *communication*, or *gestures* were not mentioned. Families were asked to continue their everyday activities and simply ignore the observer who was sitting at an appropriate distance (e.g., in the corner of the room). Families were informed that they did not need to stay inside but could pursue their daily routines or whatever they would normally do when the observer was not around (including shopping, visiting friends, picking up older siblings from school, going to the playground, playing in the garden, etc.). In those cases the observer followed at an appropriate distance but close enough to continue observation.

For the scan sampling portion of the study, each family was observed once in the morning and once on a different day in the afternoon or early evening. The observer observed a focal infant online using a pen, coding sheet with predefined categories, and a stopwatch. The video recordings were done on another day after the scan sampling was completed, once in the morning and once in the afternoon or early evening. The infant and any interlocutor (including siblings, peers, other adults; not only biological parents) were video-recorded with a hand camera at an appropriate distance. For the Dutch and the Chinese sample a second camera recorded the scene on a tripod. This arrangement did not work out in the Mayan sample because families spent most of their time outside and walked around. Two families (one Mayan and one Dutch) preferred not to be video-recorded.

Coding

Scan sampling. The sessions were divided into 5-min observation units, totaling 48 observation units. Within each unit the occurrences of predefined activities were coded. Activities were conceptualized as "states" that extended in time (Bakeman & Adamson, 1984). An activity had to last at least 30 s. Brief actions or shorter periods of fluctuating activities were not coded. Activities were coded once per observation unit (see Table 1). Of main interest was the code "triadic joint action." Triadic joint action was coded when the infant and another person engaged together about an external entity or event, for example, by acting on an object ("joint engagement"; Bakeman & Adamson, 1984). Other categories coded for "primary care" (attending to infants' biological needs), "social orientation" (observing other people), and "individual activities" (infants acting on their own; being self-absorbed). To account for the inverse relation between number and duration of activities in a unit, we divided a given activity in a unit by the sum of all other activities in that unit. We analyzed the mean proportion of time spent in each of the five activity categories across the 5-min units.

We also coded for the distance to the nearest potential interlocutor to check whether the availability and thus possibility to communicate was different across the three cultures. Distance coding was independent of the activity coding but followed the same rules. Five distances fell into three main categories: (a) *attached*: infants are attached to a caregiver's body, or infant and caregiver touch each other intentionally; (b) *around*: infant and caregiver are within reach or in view; and (c) *away*: caregivers are out of infants' view.

Reliability of the observer could only be assessed in retrospect by having the observer and a second assistant independently watch six of the video recordings (two from each culture, totaling 36 observation units) and use the same online scan sampling procedure. Agreements for the activities

Table 1
Coding Categories for Scan Sampling

Categories	Activities	Definition	Examples
Interaction	Triadic joint action	Infant and interlocutor are engaged about an external entity or event	Interlocutor and infant play with toys Interlocutor and infant look at pictures/books
	Dyadic interaction	Infant and interlocutor interact with each other without an object	Interlocutor makes funny faces for the infant Interlocutor twists infant around in the air; holds infant on her hands and walks around with her
Primary care	Washing/dressing	Infant is washed/dressed	Person washes/bathes/dresses infant Person changes infants' diaper
	Feeding	Infant is fed	Mother breastfeeds infant Person feeds infant with crackers, tortillas, etc.
Social orientation	Watching other people	Infant watches other people (without being involved in the event herself)	Infant watches mother talking to the shop assistant Infant watches other children running around
Individual activities	Object play	Infant plays with an object by herself	Infant plays with a toy Infant picks leaves from a bush
	Moving	Infant moves around by herself	Infant crawls on the floor Infant walks around in a walker
	Eating	Infant eats by herself	Infant drinks from her bottle Infant eats apple pieces
	Unengaged Other	Infant does nothing (not even watches)—just sits, lies around being self-absorbed	Infant lies in a hammock doing nothing Infant sits on caregiver's lap doing nothing

and the distance coding were excellent, respectively Cohen's Kappas = .947 ($N = 360$) and .924 ($N = 180$).

Video analyses of gestures. We coded deictic gestures produced by and for the infant. As with triadic interactions, deictic gestures involve a referential triangle of sender, recipient, and object or event. They serve the primary function of directing attention to the object or event. POINT: The arm is either fully or half extended toward an object or event (close or distant). We distinguished index-finger points (the index finger is extended relative to other fingers) from whole-hand points (no index-finger extension). SHOW: Hand holds out an object, arm is extended toward a person. PLACE: Object is placed in front of a person, within her personal space, to direct attention to it (see also Clark, 2003). Mere object transfer was not counted. OFFER: Hand holds an object that is brought close to a person, so she can take it. Interlocutors' gestures had to be directed ostensively at the infant as a recipient—gestures to third parties were not counted as direct input.

Two research assistants, naive to the purpose of our hypotheses, coded the videos with the computer-based program ELAN. They coded equal portions of the three cultural samples and an overlap of 9 infants (20%; 3 infants per culture) to establish reliability. The assistants' coding of infants' and others' gestures was highly correlated and statistically not different from each other (for infants, all $rs > .948$, $ps < .001$, $ts < 1$; for others, all $rs > .913$, $ps \leq .001$, $ts < 1$). Since data revealed great variability within cultures on several measures and diverged significantly from the normal distribution (Kolmogorov–Smirnov tests, $ps < .05$), data were subjected to square root transformations (Osborne, 2002). For ease of interpretation, the figures display uncorrected means.

Results

Scan Sampling

Triadic joint action. A 3 (culture) × 2 (age group) analysis of variance (ANOVA) confirmed our prediction. Figure 1 shows that Mayan infants spent less time in triadic interactions compared to Dutch infants, who in turn spent less time compared to Chinese infants: linear contrast, $F(2, 42) = 46.44$, $p < .0005$; planned comparisons (LSD) based on the overall variance of the ANOVA, respectively, $p = .035$; $p < .0005$. There was a main effect of age with more triadic interaction in the older age group, $F(2, 42) = 11.10$, $p = .002$, but this effect did not interact with culture.

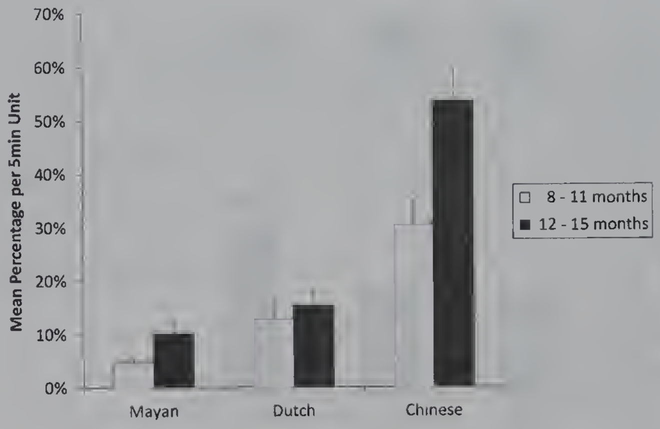


Figure 1. Triadic joint action between infant and interlocutor.

Distance. A 3 (distance) × 3 (culture) ANOVA revealed a significant main effect of distance, $F(2, 90) = 113.39$, $p < .001$, which interacted with culture, $F(4, 90) = 10.02$, $p < .001$. Simple effects (univariate) revealed that cultures varied significantly from one another in all three distances, all $Fs(2, 45) > 5.9$, $ps < .005$. Pairwise comparisons (LSD) based on the overall variance of the ANOVA showed that Dutch infants were less often *attached* compared to Mayan and Chinese infants (both $ps < .0005$), and instead more often *around* (respectively, $p = .002$; $p = .019$) or *away* (both $ps < .0005$); see Figure 2. Mayan and Chinese infants did not differ significantly from one another on either of the distances.

Other activities. Figure 3 shows the mean proportion of all activities for each of the three cultures. In all three cultures infants spent about equal time with their caregivers in primary care. A significant one-way ANOVA, $F(2, 45) = 18.0$, $p < .0005$, revealed that Chinese infants spent less time in individual activities than Mayan and Dutch infants, presumably reflecting Chinese caregivers' role in creating triadic play episodes (post hoc comparisons [LSD] based on the overall variance, both

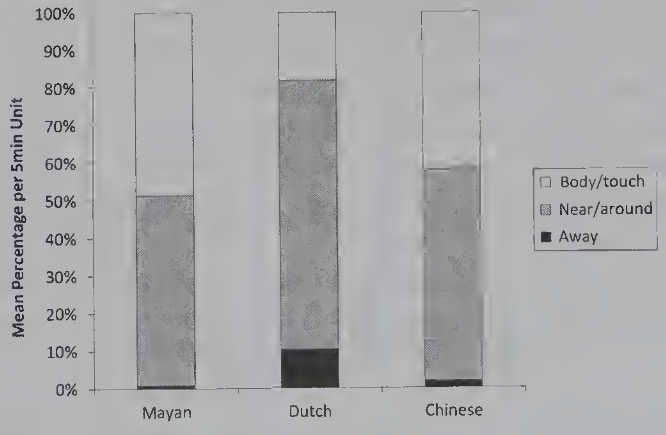


Figure 2. Interlocutor distance.

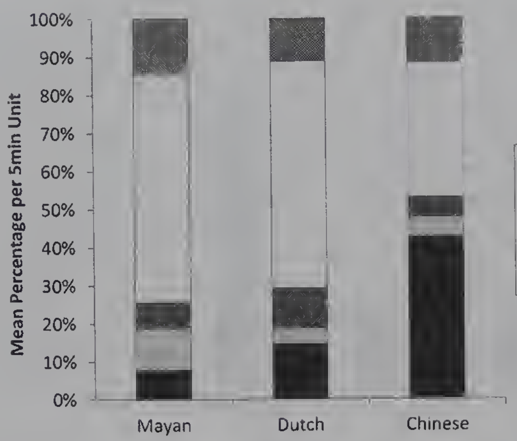


Figure 3. Daily activities.

$ps < .0005$). A further significant one-way ANOVA, $F(2, 45) = 5.59$, $p < .007$, revealed that Mayan infants spent more time in dyadic interaction with their caregiver than Dutch and Chinese dyads (post hoc comparisons [LSD] based on the overall variance respectively, $p = .028$, $p = .002$). Finally, a significant one-way ANOVA, $F(2, 45) = 4.4$, $p < .018$, revealed that Dutch infants spent slightly more time watching other people than Mayan or Chinese infants (post hoc comparisons [LSD] based on the overall variance respectively, $p = .041$; $p = .006$).

Video Analyses of Gestures

Interlocutors. A 3 (culture) × 2 (age group) ANOVA confirmed our prediction. The upper left panel of Figure 4 shows that Mayan interlocutors gestured less compared to Dutch interlocutors, who in turn gestured less compared to Chinese interlocutors: linear contrast, $F(2, 40) = 15.94$, $p < .0005$; planned comparisons [LSD] based on the overall variance of the ANOVA, respectively, $p = .050$; $p < .0005$. There was a main effect of age with more gestures to infants in the older age group, $F(1, 40) = 5.14$, $p = .029$. Age and culture interacted only marginally, $F(2, 40) = 2.89$, $p = .069$. Visual inspection of the data suggests an absence of age differences in the Dutch but not in the Mayan and Chinese samples.

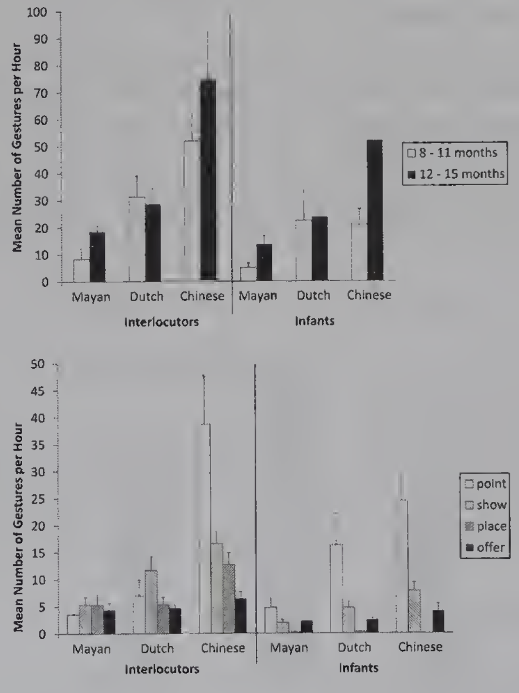


Figure 4. Upper panel: interlocutors' and infants' gestures. Lower panel: gesture types used by interlocutors and infants.

The lower left panel of Figure 4 shows the types of interlocutors' gestures directed to infants across cultures. A 4 (gesture) × 3 (culture) ANOVA revealed main effects of gesture, $F(3, 129) = 5.19$, $p = .002$, and culture, $F(2, 43) = 19.95$, $p < .001$, which were qualified by an interaction, $F(6, 129) = 6.49$, $p < .001$. Simple effects (univariate) revealed that cultures varied significantly from one another in all but the OFFER gesture, all $Fs(2, 43) > 6.7$, $ps < .003$. For POINT, SHOW, and PLACE, pairwise comparisons (LSD) based on the overall variance revealed that Chinese interlocutors used POINT and PLACE significantly more often than Mayans and Dutch (all $ps < .003$). Chinese and Dutch used SHOW significantly more often than Mayans ($ps < .017$), and Chinese tended to use SHOW more often than Dutch ($p = .090$).

Within a culture, simple effects (multivariate) revealed that gesture types differed significantly from one another only in the Dutch and Chinese samples, respectively, $F(3, 41) = 4.3$, $p = .010$; $F(3, 41) = 15.94$, $p < .0005$. Pairwise comparisons (LSD) based on the variance of the overall ANOVA revealed that Chinese interlocutors used POINT more frequently than SHOW ($p = .011$), SHOW and PLACE about equally often, and OFFER less than all other gestures ($ps < .009$). Dutch interlocutors used SHOW more than all other gestures ($ps < .013$), with no significant differences between the other gestures. Mayan interlocutors used all gestures with equal frequency.

Infants. A 3 (culture) × 2 (age group) ANOVA confirmed our prediction. As shown in the upper right panel of Figure 4, Mayan infants gestured less compared to Dutch infants, who in turn gestured less compared to Chinese infants: linear contrast, $F(2, 40) = 9.81$, $p = .001$; planned comparisons (LSD) based on the overall variance of the ANOVA, respectively, $p = .045$; $p = .029$. There was a main effect of age with more gestures in the older age group, $F(1, 40) = 11.61$, $p = .002$, but this effect did not interact with culture.

The lower right panel of Figure 4 shows the types of infant gestures across cultures. A 4 (gesture) × 3 (culture) ANOVA revealed main effects of gesture, $F(3, 129) = 47.5$, $p < .001$, and culture, $F(2, 43) = 5.78$, $p = .006$, which were qualified by an interaction, $F(6, 129) = 4.98$, $p = .001$. Simple effects (univariate) revealed that cultures differed significantly from one another for POINT, $F(2, 43) = 7.01$, $p = .002$, and SHOW, $F(2, 43) = 5.21$, $p = .009$, but not for OFFER and PLACE. Pairwise comparisons (LSD) based on the variance of the overall ANOVA confirmed for POINT the pattern of Mayan < Dutch < Chinese (respectively, $p = .030$, one-tailed; $p = .040$, one-tailed). For SHOW the pattern was similar, Mayan < Dutch ≤ Chinese (respectively, $p = .042$, one-tailed; $p = .081$, one-tailed).

Within a culture, simple effects (multivariate) revealed that gesture types differed significantly from one another in each of the three cultures, all three $Fs(3, 41) > 5.14$, $ps < .004$. Pairwise comparisons (LSD) based on the variance of the overall ANOVA revealed that this was the case because PLACE was used in all three cultures significantly less compared to all other gestures (i.e., hardly or never; all $ps < .013$). Chinese and Dutch infants used POINT more often than SHOW (respectively, $p < .0005$; $p = .030$), and SHOW more often than OFFER (respectively, $p = .003$; $p = .036$). In contrast, Mayan infants used these gestures with equal frequency.

Figure 5 shows that only 4 of the Mayan infants (27%) pointed with the index finger (with a mean number of points = 6.8). This was in contrast to 11 Dutch infants (72%; mean number of points = 14.5) and 14 Chinese infants (88%; mean number of points = 17.0). The number of index-finger pointers was significantly different across the three groups, $\chi^2(2, 46) = 13.3$, $p = .001$. Importantly, the 4 Mayan infants who pointed with the index finger were in the older age group above 12 months of age. Conversely, the 4 Dutch and 2 Chinese infants who did not yet point with the index finger were in the younger age group and all at or below 10 months of age.

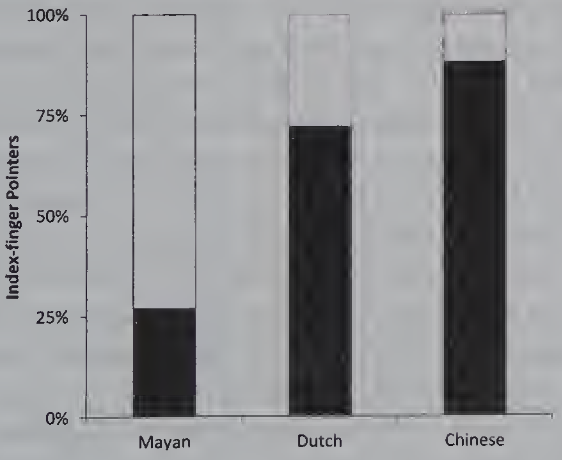


Figure 5. Percentage of index-finger pointers.

Predictors of Infant Gestures

In a final analysis we tested whether differences in infants' gestures could be determined by positive correlations with interlocutors' gestures and triadic social interactions, even when controlling for culture. There were substantial correlations between interlocutors' gestures and infants' gestures, $r(df = 46) = .626$, $p < .0005$, and between triadic social interactions and infants' gestures, $r(df = 46) = .668$, $p < .0005$. Positive partial correlations remained when controlling for culture (dummy-coded), respectively, $r(\text{par}, df = 42) = .414$, $p = .005$; $r(\text{par}, df = 42) = .480$, $p = .001$, and when controlling for both culture and age (continuous), $r(\text{par}, df = 41) = .273$, $p = .038$, one-tailed; $r(\text{par}, df = 41) = .251$, $p = .052$, one-tailed.

In addition, partial correlations revealed that interlocutors' gestures and triadic social interactions were interrelated, $r(\text{par}, df = 41) = .405$, $p = .007$. We investigated whether this interrelation differentially affected the correlations with infants' gestures. Infants' gestures did not correlate significantly with interlocutors' gestures when controlling for triadic social interactions, $r(\text{par}, df = 43) = .215$, $p = .165$. However, infants' gestures remained correlated with triadic social interactions when controlling for interlocutors' gestures, $r(\text{par}, df = 43) = .364$, $p = .014$. These findings thus suggest a more direct link between triadic social interaction and infants' gestures.

Discussion

On- and off-line observations of infants' engagement in joint actions and gestural communication revealed commonalities and differences across very different cultural settings. Despite vast sample differences, for example, in societal structure, parental beliefs, economy, educational and socialization goals, geographical and physical layout, daily structure and family size, we observed in all cultural settings that infants were engaged in triadic bouts of joint action, that they were gestured for, and that they gestured themselves. All four types of deictic gestures were present in the very different cultural settings. These findings support the assumption that joint action and deictic gestures are universally existing prelinguistic behaviors, and corroborate recent findings of a prelinguistic gestural universal of human communication in the specific case of index-finger pointing (Liszkowski et al., 2012).

The current study asked whether differences in the extent to which infants are immersed in social interaction and gestural communication would impact infants' use of gestures. Results revealed systematic differences between sociocultural settings in the frequency with which interlocutors engaged infants in joint actions and gestural communication. Our method of natural observations captured these real-life differences better than a previous study (Liszkowski et al., 2012) which used a prestructured elicitation format and probed only for the existence of a behavior within a given context, without controlling for its natural occurrence or cultural relevance.

The cultural group differences in the social–interactional "input" to infants did not simply result from interlocutors' proximity. Although Yucatec Mayans and Shanghai Chinese were on the opposite extremes of our social interaction measures, they did not differ in caregiver proximity: Infants were equally often at or around caregivers and available for interaction. Thus, we can exclude the possibility that proximity selectively prevented (or enhanced) joint actions and gestures. The differences also did not result from differences in mobility constraints. Mayan infants were not constrained to interact by being swaddled or carried in tight shawls. Instead, they could freely move most of the time. The number of potentially available interlocutors also cannot account for the differences we observed. For example, Mayan infants usually encountered more people as they spent a lot of time outside compared to the Chinese infants who stayed often (but not exclusively) inside small flats. Also, Dutch infants were more often alone than infants in the other two samples, presumably due to the use of playpen. Further, like Chinese, most Dutch infants in the current study were firstborn, suggesting that birth order itself was not the only driving factor of the differences we observed. And although socioeconomic status (SES) and education have been associated with caretaking patterns (Richman, Miller, & LeVine, 1992), our post hoc judgment is that the Shanghai parents did not seem to have a higher education or SES than the Dutch parents. Note also that the relative meaning of education and SES within a cultural setting may be difficult to compare across cultures. Importantly, the cultural samples were not different altogether, as we controlled for primary caretaking activities that were similar across settings. Thus, the current study revealed systematic quantitative differences in the amount of triadic interactions and deictic gestures to which young infants are exposed. These differences appeared on the level of broadly defined cultural settings (Whiting, 1980), which constitute the developmental and ecocultural niches within which infants develop (see Super & Harkness, 1986; Weisner, 2002).

Having established quantitative group differences in the interactional input, this allowed us to test the hypothesis that social–interactional input influences infants' prelinguistic gestures from early on. The natural input differences to age-matched infants of different cultural groups provided quasi-experimental independent conditions for a direct test. This is a lucky situation as the experimental suppression of joint acting and communicating with infants within a cultural setting is difficult, if not impossible, for ethical reasons. As predicted by the hypothesis, infants' prelinguistic gestures differed across cultural settings in the same way as their social--interactional input. Dutch infants who were exposed to more interactional input than Mayan infants gestured more, yet Chinese infants who were exposed to more interactional input than Dutch infants gestured still more. The specific pattern across three cultural samples makes it unlikely that any specific characteristic of one cultural sample alone accounted for these differences. Further, all infants were age-matched across the three cultural samples, suggesting that cultural differences impact infants' gesture use from early on.

The analysis of index-finger pointing supports the interpretation on the individual level. Index-finger pointing has previously been shown to be a milestone in communicative and social-cognitive development (Liszkowski & Tomasello, 2011; Rowe & Goldin-Meadow, 2009). In the current study, index-finger pointing was more frequent than other gestures and accounted for most of the group differences in infants' gestures. On an individual level, while all infants gestured at least once (except for two of the youngest infants in each of the Dutch and the Yucatec Mayan samples), index-finger pointing had emerged in only a few, older Mayan infants, while it had already emerged in the majority of the Dutch, and almost all Chinese infants. Research has revealed a correlated usage of parent and infant pointing across various cultural settings (Liszkowski & Tomasello, 2011; Liszkowski et al., 2012), and this relation presumably mediates later word acquisition and preschooler's vocabulary size (Rowe & Goldin-Meadow, 2009). The current finding on the individual level suggests that the very emergence of this developmental milestone is itself mediated by the amount of earlier social–interactional experience.

Our correlation analyses further revealed that the relation between social–interactional input and infants' gesture use was not mediated by culture, suggesting that social–interactional input is a universal developmental mechanism in the emergence of prelinguistic communication. When looking at the contributions of joint action and deictic gestures separately, the set of partial correlations revealed that infants' gestures were not predicted by interlocutors' gestures once we controlled for the time spent in joint action. The analyses on the frequency of the different gesture types further suggest that there was no direct one-to-one correspondence in the forms of interlocutors and infants' gestures. For example, as displayed in Figure 4, Dutch infants were exposed to more SHOW gestures but used more POINT than SHOW gestures; and although infants were exposed to the PLACE gesture they hardly ever used it. Presumably, deictic gestures do not emerge simply by mimicry alone. Instead, infants' gesture use may in part be mediated by a form-independent use of reference in the sense that acting on objects brings these to the attention of others. Indeed, infants' gestures were predicted by the time spent in joint action, and the relation remained when controlling for the amount of gestures directed at them. Word learning studies have shown that time spent in joint engagement is predictive of vocabulary size (Tomasello & Farrar, 1986). Based on the current findings it is likely that not only verbal reference, but even the very concept of prelinguistic reference itself emerges within shared activities, in which others' actions structure infants' attention to objects.

An alternative possibility would have been that socialization processes become effective only after infants participate actively in social exchanges. In this case one would have expected no cultural differences early in infancy. This alternative was not supported. Yet, as a cross-cultural approach is only quasi-experimental in that there is never full control over the independent variable, it is at least theoretically possible that a nonsocial factor (e.g., one under genetic control) could account for the pattern of group differences. However, to date it is unknown whether there is relevant genetic variation in the expression of infant gestures, and we would find it unlikely that this variation would pattern across the three cultures in just the same way our behavioral findings did. A Gene × Environment interaction, as for example in the case of linguistic tone (Dediu, 2011), remains of course a possible, albeit to date speculative, interpretation of the current findings. Note also that while nutritional deficiencies may have an effect on early development (see Gorman, 1995), there was no known history of malnutrition in our samples, and certainly not in a way plausibly explaining the obtained differences between all three cultures. Instead, the current findings are consistent with the interpretation that the interactional input influences the very emergence of infants' prelinguistic gestures.

It is also unlikely that the 1 hr of video recording was too short to capture instances of index-finger pointing in some samples. First, the observation interval was equal in all three cultures. Second, the observation interval was long enough to yield ample use of other object-directed gestures. Third, infants who did point with the index finger clearly pointed more than once per hour (also in the Mayan sample). It is also implausible that differences in the amount of items to point at, especially toys, drove the differences. First, the Yucatec Mayan infants who did not point used other, object-directed gestures, which must mean that there were objects to gesture with. Second, pointing increased with age, while the environment remained the same, again showing that there were items to point at, once the behavior had emerged.

The current study did not distinguish between the finer grained, mostly conceptual distinction of "active" and "passive" joint engagement, that is, respectively, how many of the joint bouts were initiated and maintained by the infant, and how many were scaffolded by the interlocutor. However, Bakeman and Adamson (1984) found that most of the shared activities in that age range are still scaffolded by caregivers, which also corresponds to our anecdotal observations. Another aspect that might warrant further investigation is whether joint actions differ also qualitatively across cultures, and whether certain types of joint activity will be more—or less—inductive to infant gesturing (see Puccini, Hassemer, Salomo, & Liszkowski, 2010). Note also that we did not obtain a clear indication of "optimal" input: While Dutch infants, taken as a reference sample to most developmental literature, received more input than Yucatec Mayan infants, they received much less input than the Shanghai Chinese infants.

Explaining the origins of interlocutors' group differences touches upon the deeper, historical origins of cultural differences. Perhaps it is less likely that there is a direct link to the independent–interdependent socialization goals dichotomy that is often invoked in cross-cultural research (e.g., Keller, 2007), as both Yucatec Mayans and Shanghai Chinese have characteristics of following interdependent socialization goals (albeit in rural vs. urban settings; cf. Kagitcibasi, 2005). More directly, culturally transmitted beliefs about infants and development, environmental factors and habitual practices that frame the business of daily life, education, as well as available cultural tools, presumably all affect the style and frequency of social interaction with nonverbal infants (e.g., Chisholm, 1981; Harkness & Super, 1996; Richman et al., 1992).

Development takes place within sociocultural activity. To benefit from and participate in sociocultural activities, one needs some fundamental communication skills. The current study reveals that the emergence of prelinguistic gestures is already mediated by social–interactional experience. The analyses of natural input differences across different sociocultural settings provided quasi-experimental evidence for a socially mediated emergence of prelinguistic deictic gestures in infancy. Our findings thus support sociocultural accounts of development already before language or other symbolic tools are in play. It is a reasonable assumption that others' actions structure the ontogeny of human communication and social cognition from early on.

References

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother–infant and peer–infant interaction. *Child Development*, *55*, 1278–1289.

Brown, P. (2000). Conversational structure and language acquisition: The role of repetition in Tzeltal. *Journal of Linguistic Anthropology*, *8*, 197–221.

Carpendale, J. I. M., & Carpendale, A. B. (2010). The development of pointing: From personal directedness to interpersonal direction. *Human Development*, *53*, 110–126.

Carpendale, J., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*, 79–96.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4, Serial No. 255).

Chisholm, J. S. (1981). Residence patterns and the environment of mother–infant interactions among the Navajo. In T. Field (Ed.), *Culture and early interactions* (pp. 3–19). Hillsdale, NJ: Erlbaum.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243–268). Mahwah, NJ: Erlbaum

Dediu, D. (2011). A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society of London B*, *278*, 474–479.

Feiring, C., & Lewis, M. (1981). Middle class differences in the mother–child interaction and the child's cognitive development. In T. Field (Ed.), *Culture and early interactions* (pp. 63–90). Hillsdale, NJ: Erlbaum

Fong, V. L. (2007). Parent–child communication problems and the perceived inadequacies of Chinese only children. *Ethos*, *35*, 85–127.

Gaskins, S. (1996). How Mayan parental theories come into play. In S. Harkness & C. M. Super (Eds.), *Parents' cultural belief systems: Their origins, expressions, and consequences* (pp. 345–363). New York: Guilford

Gaskins, S. (1999). Children's daily lives in a Mayan village: A case study of culturally constructed roles and activities. In A. Göncü (Ed.), *Children's engagement in the world* (pp. 25–81). Cambridge, UK: Cambridge University Press.

Gaskins, S. (2000). Children's daily activities in a Mayan village: A culturally grounded description. *Journal of Cross-Cultural Research*, *34*, 375–389.

Gaskins, S. (2006). Cultural perspectives on infant–caregiver interaction. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality* (pp. 279–298). Oxford, UK: Berg.

Goh, E. C. L. (2006). Raising the precious single child in urban China: An intergenerational joint mission between parents and grandparents. *Journal of International Relationships*, *4*, 7–28.

Goh, E. C. L., & Kuczynski, L. (2010). Agency and power of single children in multi-generational families in urban Xiamen, China. *Culture & Psychology*, *15*, 506–532.

Goldin-Meadow, S., & Saltzman, J. (2000). The cultural bounds of maternal accommodation: How Chinese and American mothers communicate with deaf and hearing children. *Psychological Science*, *11*, 307–314.

Gorman, K. S. (1995). Malnutrition and cognitive development: Evidence from experimental/quasi-experimental studies among the mild-to-moderately malnourished. *Journal of Nutrition*, *125*, 2239S–2244S.

Harkness, S., & Super, C. M. (Eds.). (1996). *Parents' cultural belief systems: Their origins, expressions, and consequences*. New York: Guilford.

Harkness, S., Super, C. M., & van Tijen, N. (2000). Individualism and the "Western mind" reconsidered: American and Dutch parents' ethnotheories of the child. *New Directions for Child and Adolescent Development*, *87*, 23–39.

Harris, M., Barlow-Brown, F., & Chasin, J. (1995). The emergence of referential understanding: Pointing and the comprehension of object names. *First Language*, *15*, 19–34.

Hewlett, B. S. (1996). *Diverse contexts of human infancy*. New York: Prentice Hall.

Ho, D. (1989). Continuity and variation in Chinese patterns of socialization. *Journal of Marriage and Family*, *51*, 149–163.

Iverson, J., & Goldin-Meadow, S. (2005). Gestures paves the way for language development. *Psychological Science, 16*, 367–371.

Kagitcibasi, C. (2005). Autonomy and relatedness in cultural context. *Journal of Cross-Cultural Psychology, 36*, 403–422.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Erlbaum.

Keller, H., Otto, H., Lamm, B., Yovsi, R., & Kärtner, J. (2008). The timing of verbal/vocal communication between mothers and their infants: A longitudinal cross-cultural comparison. *Infant Behavior & Development, 31*, 217–226.

Kun, Z. (2007). 50% plus of city's under-3s raised by grandparents. *China Daily*, 13 July, p. 5.

Lancy, F. (2010). Learning 'From nobody': The limited role of teaching in folk models of children's development. *Childhood in the Past: An International Journal, 3*, 79–106.

Lieven, E., & Stoll, S. (2009). Language development. In M. Bornstein (Ed.), *The handbook of cross-cultural developmental science* (pp. 143–160). New York: Psychology Press.

Liszkowski, U. (2010). Deictic and other gestures in infancy. *Acción Psicológica, 7*, 21–33.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science. 36*, 698–713. doi:10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development, 26*, 16–29.

Osborne, J. (2002). Notes on the use of data transformation. *Practical Assessment, Research & Evaluation, 8*(6). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=6

Puccini, D., Hassemer, M., Salomo, D., & Liszkowski, U. (2010). The type of shared activity shapes caregiver and infant communication. *Gesture, 10*, 279–297.

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language, 13*, 85–100.

Pye, C. (1992). The acquisition of K'iche' Maya. In D. Slobin (Ed.), *The cross-linguistic study of language acquisition*, Vol. 3 (pp. 221–308). Hillsdale, NJ: Erlbaum.

Richman, A., Miller, P., & LeVine, R. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology, 28*, 614–621.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*, 951–953.

Schieffelin, B., & Ochs, E. (Eds.) (1986). *Language socialization across cultures*. New York: Cambridge University Press.

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development, 9*, 545–569.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*, 1454–1463.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*, 197–212.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weisner, T. S. (2002). Ecocultural understanding of children's developmental pathways. *Human Development, 45*, 275–281.

Weisner, T. S. (2009). Culture, development, and diversity: Expectable pluralism, conflict, and similarity. *Ethos, 37*, 181–196.

Whiting, B. (1980). Culture and social behavior: A model for the development of social behavior. *Ethos, 8*, 95–116.

Whiting, B., & Whiting, J. (1975). *Children of six cultures: A psycho-cultural analysis*. Cambridge, MA: Harvard University Press.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1308–1323

Dual Routes to Cognitive Flexibility: Learning and Response-Conflict Resolution in the Dimensional Change Card Sort Task

Michael Ramscar
University of Tübingen

Melody Dye
Indiana University

Jessica W. Gustafson and Joseph Klein
Stanford University

Cognitive control, the ability to align our actions with goals or context, is largely absent in children under four. How then are preschoolers able to tailor their behavior to best match the situation? Learning may provide an alternative route to context-sensitive responding. This study investigated this hypothesis in the Dimensional Change Card Sort (DCCS), a classic test of cognitive control that most under-fours fail. A training intervention based on learning theoretic principles proved highly effective: Three-year-olds who learned about DCCS rules and game contexts in a card-labeling task, subsequently transferred this knowledge to sorting in the DCCS, passing at more than 3 times the rate of controls ($N = 47$). This surprising finding reveals much about the nature of the developing mind.

Although a 3-year-old girl might appear to simply be a smaller version of her 4-year-old brother, studies have revealed some surprising differences between them: While her older brother will successfully pass most of the tasks that have been devised to test the cognitive capabilities of the very young, odds are that she will fail every one of them. At test, a typical 4-year-old can ably navigate the conflicting dimensions of appearance and reality, correctly judge questions of false belief, successfully switch between competing rules in the Dimensional Change Card Sort (DCCS), and when faced with a forced choice task, reliably choose between two alternatives according to their prior probabilities. An average 3-year-old, on the other hand, will maximize in that same task (fixating on the most likely response), fail to distinguish reality from false belief and appearance from reality in standard batteries, and fail to switch from one sorting rule to another in the DCCS, even when the new rule is clearly stated (for reviews, see Diamond, 2002; Hanania & Smith, 2009; Ramscar & Gitcho, 2007).

Over the past two decades, the underlying causes of these phenomena have been much debated, and many conflicting explanations have been offered to account for 3-year-olds' curious performance in these tasks (see, e.g., cognitive control & complexity theory, Zelazo, Müller, Frye, & Marcovitch, 2003; active vs. latent representations, Cepeda & Munakata, 2007; Diamond, Carlson, & Beck, 2005; Morton & Munakata, 2002; Yerys & Munakata, 2006; attentional inertia, Kirkham, Cruess, & Diamond, 2003; redescription, Kloo & Perner, 2005; Kloo, Perner, Kerschhuber, Dabernig, & Aichhorn, 2008; negative priming, Müller, Dick, Gela, Overton, & Zelazo, 2006; Perner, Stummer, & Lang, 1999). While the bulk of these proposals have been aimed at accounting for why 3- and 4-year-olds perform so differently, an equally puzzling question has been accorded rather less attention: Given the inflexibility of thought revealed in these tests, why is this rigidity so difficult to detect in the normal course of events? Why is it that young children appear to be capable of responding in flexible and context sensitive ways in some situations, but not in others (Brooks, Hanauer, Padowska, & Rosman, 2003; Deák, 2003)?

Here, we outline a solution to this puzzle, based on a consideration of the different ways in which a

---

This material is based on work supported by the National Science Foundation under Grants 0547775 and 0624345 and by the National Science Foundation/IGERT Training Program in the Dynamics of Brain-Body-Environment Systems at Indiana University. We are grateful to Natasha Kirkham, Sharon Thompson-Schill, and Stewart McCauley for discussion of these ideas.

Correspondence concerning this article should be addressed to Michael Ramscar, Department of Linguistics, Eberhard Karls University, Tübingen, Germany 72074. Electronic mail may be sent to michael.ramscar@uni-tuebingen.de.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0015
DOI: 10.1111/cdev.12044

child might resolve the conflict between competing ways of responding to a given situation, so as to choose the response most appropriate to that context. We propose that there are at least two ways in which response conflict can be handled in the mind: dynamic response conflict resolution, which processes and resolves conflicting response demands online, as they arise, and contextual discrimination learning, which shapes the degree to which responses are evoked in specific contexts and which can serve to reduce (or even eliminate) potential response conflicts that would otherwise need to be resolved online. While older children and adults will in most cases have access to both "routes" of response-conflict resolution, we suggest that children under four are generally limited to one: *learning*. Under-fours are able to match their behavior to context in remarkably subtle and sensitive ways, but only once they have learned to do so. If they have not learned to match a response or a behavior to a context, under-fours' difficulty with handling online response conflict proves their undoing in psychological tests.

Response Conflict and the DCCS Task

In the standard DCCS, children are asked to sort cards with two prominent dimensions—color and shape—into bins in which these dimensions have been reversed (Figure 1). For example, if a child is holding cards with red stars and blue trucks, the bins will be marked with blue stars and red trucks. If the child is asked to sort by color, the red stars will go with the red trucks and the blue stars will go with the blue trucks; if the child is asked to sort by shape, the red stars will go with the blue stars, and the red trucks will go with the blue trucks. Importantly, when a child is asked to sort by one dimension—say, shape—changing the sort dimension to color will also change the correct sort bin: For example, red stars go in the truck bin when sorted by color, but the star bin when sorted by shape. Successful sorting in the postswitch thus requires the child to actively choose between conflicting responses, suppressing the dominant sorting response (which has been strongly associated with the cards over the initial game trials) in favor of a contradictory alternative (which is now relevant to the task at hand).

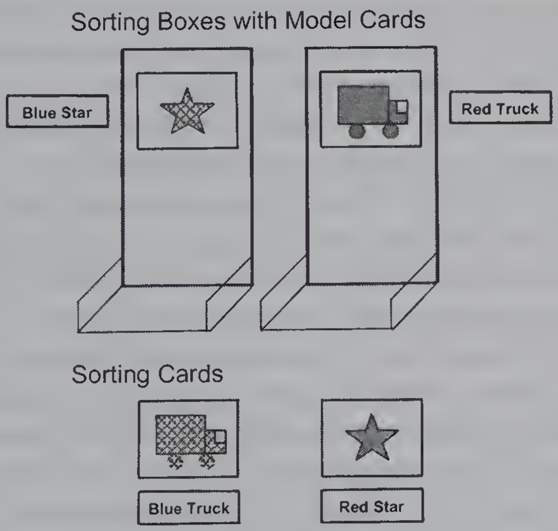


Figure 1. The Dimensional Change Card Sort (DCCS) task. Cards can be sorted by *shape* (in which case, the red star card is sorted into the left bin) or *color* (in which case, the red star card is sorted to the right bin).

Response conflict will arise whenever the requirements of a specific task conflict with an equally or more strongly learned pattern of responding that is prompted by the same context. To successfully resolve this conflict, an individual must be able to effectively override the biased response in favor of a less well-learned (or less well-primed) response that is more appropriate to the context (Novick, Trueswell, & Thompson-Schill, 2010). In adults, the ability to actively and selectively respond in this way can be easily illustrated in relation to the Stroop test, in which subjects are asked to identify the ink color of a conflicting color word. For example, if *green* is printed in red ink, *red* is the appropriate response. Success in this task involves resolving the conflict between the dominant response (reading) and a disfavored but contextually appropriate response (ink naming). Although the undertaking is cognitively demanding (especially when a time component is introduced), most adults are able to pass the Stroop if they are given time to formulate their response.

The DCCS imposes a similar set of task demands, but in a game designed for children. To succeed in the DCCS, a child must be able to resolve response conflict online as the sorting requirements change between games. Unlike adults and older children, however, 3-year-olds not only tend to fail the DCCS but also the broader battery of behavioral measures designed to test various aspects of online response selection (Ramscar & Gitcho, 2007). For the DCCS at least, this poor performance is unlikely to be attributable to the verbal sophistication of the task: Three-year-olds can correctly answer questions about how to sort cards according to the game rules, even as the rules change, and consistently sort correctly in the first game (Zelazo, Frye, & Rapus, 1996; but see Munakata & Yerys, 2001). It is not until the beginning of the second game, when the child is handed a fresh card and asked to sort by the new rule that the difficulty emerges: Three-year-olds who appear to "know" that the rules have changed nevertheless fail to switch their responses, and continue to sort by the first rule. Children make identical errors when observing a puppet sort the cards (Jacques, Zelazo, Kirkham, & Semcesen, 1999), suggesting that the difficulty is not solely one of motor perseveration. This pattern of responding is seen up until around age 4, when children begin to consistently pass the DCCS, successfully matching the cards to the bins both before and *after* the sorting rules are changed (Zelazo et al., 2003).

Why do 3-year-olds typically fail the DCCS, while 4-year-olds pass it? What aspects of the underlying developmental process motivate this transition? One promising suggestion is that 3-year-olds' poor performance is related to the immaturity of the prefrontal cortex (PFC; Brooks et al., 2003; Deák, 2003; Diamond, 2002; Moriguchi & Hiraki, 2009; Ramscar & Gitcho, 2007). PFC has long been implicated in flexible, goal-directed behavior (Miller & Cohen, 2001), and collectively, its regions are thought to act as a "dynamic filter" that selectively maintains appropriate patterns of neural activation, while gating inappropriate or irrelevant ones (Shimamura, 2000). Moreover, in adults, PFC is known to be a critical component in a functional circuit that appears to deal with the demands of online response-conflict processing (Yeung, Botvinick, & Cohen, 2004; Yeung, Nystrom, Aronson, & Cohen, 2006).

There is strong evidence supporting the idea that prefrontal function is late in developing (Davidson, Amso, Cruess Anderson, & Diamond, 2006) and that the neural circuitry that underpins it follows a protracted, nonlinear developmental trajectory (Ramscar & Gitcho, 2007). For example, while synaptogenesis in visual and auditory cortex peaks a few months after birth, the same developments occur only much later in PFC (Huttenlocher & Dabholkar, 1997). In additional, Von Economo neurons, which appear to play important role in signaling between the areas involved in response-conflict processing, are largely absent in newborns and only reach adult levels in the 4th year (Allman, Watson, Tetreault, & Hakeem, 2005; Ramscar & Gitcho, 2007; Rueda, Rothbart, McCandliss, Saccamanno, & Posner, 2005).

As a consequence of this prolonged period of cortical immaturity, children under four exhibit patterns of impaired behavioral and cognitive control that is similar in many ways to that of patients with neurological PFC damage (Thompson-Schill, Ramscar, & Evangelia, 2009). Young children seem largely unable to act in ways that conflict with dominant prepotent responses and are highly limited in their ability to dynamically filter their behavior or their learning in accordance with the demands of a particular context or goal (Ramscar & Gitcho, 2007). On balance then, the neurological and behavioral evidence suggests that 3-year-olds' difficulties with rule switching in the DCCS stems from the slow functional development of the neural circuitry that underlies response-conflict processing.

Discrimination Learning

While under-fours appear largely incapable of resolving conflict online, there is still another "route" to success available to them in the DCCS: learning. Children can learn that the games associated with each sorting rule provide the contextual information needed to make the appropriate response and then use this (implicit) knowledge to effectively minimize response conflict that the DCCS usually evokes, thereby allowing them to pass the task. This suggestion is supported by the finding that typical patterns of prefrontal activation seen in the structurally similar Stroop task vanish when adult participants engage in the task for a sufficient amount of time, indicating that participants learn context appropriate representations that eliminate the local response conflict that they ordinarily encounter during the course of the task (Erickson et al., 2004; see also Ramscar & Yarlett, 2007; Roediger & Karpicke, 2006). However, as children typically fail the task despite the presence of contextual cues, it also seems clear that in ordinary testing circumstances, the game cues do not provide enough scaffolding to enable 3-year-olds to pass the DCCS.

To explain why these contextual cues might still matter, we first need to consider the way that responses that lead to response conflict in the DCCS could be learned and discriminated in an implicit learning paradigm. In contemporary learning theory, learning is conceptualized as a process of discriminating ever better predictive cues to the relevant events in (or *responses to*) one's environment. Learning models embody the ideas: (a) that learning is driven by "surprise" (prediction error)—which occurs whenever our expectations and reality fail to align—and (b) that learning is attenuated when expectation and reality are aligned, such that there is little remaining uncertainty about the events that will unfold in a given context (for a review, see Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). Learning serves to reduce uncertainty about how to respond in a given context, by discriminating in favor of more informative cues to responses, and discriminating against less informative cues (i.e., cues which would lead to an inappropriate response; see Rescorla, 1988). As uncertainty diminishes, and a learner's certainty about the likelihood of a given response increases in kind, learning asymptotes.

For example, in the Rescorla and Wagner (1972) learning rule, changes in the value of cues in learning are determined by a discrepancy function ($\lambda - V$), which takes the difference between expectation and reality across trials, and uses this to update expectations by modifying the associative values of the set of cues to a given response. Cue values are strengthened when an appropriate response is underpredicted, and weakened when it is overpredicted (or predicted erroneously; Kamin, 1969; Rescorla & Wagner, 1972). Importantly, associative value lost by one cue can be subsumed by other cues, resulting in a competition for relevance between cues. The outcome of this competition is determined both by positive evidence about co-occurrences between cues and predicted events (reinforcement), and negative evidence about *non*occurrences of predicted responses (error).

Accordingly, the distribution of error associated with the *set* of cues to a particular outcome critically affects the values that individual cues take on in learning. The value of each cue is shaped by the *combined* value of both the positive and negative evidence provided by that cue, *and* the value of the evidence supplied by other *competing* cues. This results in patterns of learning that are very different from those that would result if learning were limited to simple co-occurrence tracking (a common misconception of learning; see Rescorla, 1988). Competitive discrimination learning uncovers the predictive structure of the environment, allowing learners to discover the cues that best predict appropriate outcomes (Ramscar et al., 2010; Smith, Colunga, & Yoshida, 2010).

It is worth adding that while much of the impetus for the development of these models came from behavioral experiments in animals, there is now good evidence for their neurobiological basis (see Schultz, 2006, for a review). The learning rule we have described accurately predicts patterns of synaptic firing in midbrain dopamine neurons in learning tasks (Waelti, Dickinson, & Schultz, 2001) and has been productively applied to many aspects of human behavior and cognition, such as decision making, executive function, habitual learning, and response selection (Montague, Hyman, & Cohen, 2004), demonstrating considerable predictive and explanatory power.

Method

Discriminative learning offers a mechanism that can allow young children to learn to match specific responses to specific contexts. A child can learn to better respond in context by implicitly learning to discriminate the cues to an appropriate response from unhelpful or distracting cues by monitoring the error associated with them. This process will invariably highlight contextual cues (because these will be most informative about context-appropriate responses) and devalue many context-independent cues (which may provide support for inappropriate responses). In principle, the outcome of this process should be better tuned, more context-sensitive responding.

In the DCCS, the "games" associated with each sorting rule offer just the kind of contextual information children need to help them succeed at the task. If a child was to learn to attend to (assign high values to) Conjunctive cues that take into account both the game context and the relevant sort dimension (e.g., *star* + *shape game* and *red* + *color game*), and to ignore (assign low values to) any inappropriate cues, then switching ought not to be a problem. However, to learn *what* the correct cues are, children will first have to discriminate them from cues that are less relevant to this context but that nevertheless carry valuable information, either in other contexts or across contexts: in particular, the dimensional cues, color (*red*) and shape (*star*). In what follows, we present a computational model that shows formally how perseverative sorting on switch trials is a natural result of cue competition in the early parts of this process. Before detailing how this model works, however, it is worth considering the conceptual picture that underpins our simulations of learning in the DCCS, in which the process of learning the game rules is seen as a one of tuning the expectations that "sculpt attentional biases" (Colunga & Smith, 2008, p. 180).

Imagine a 3-year-old watching her brother complete the DCCS. Although she has never played before, her brother has played many times, and he always sorts correctly. As she watches him sort cards, the girl (or, more accurately, her implicit learning mechanisms, which process information in the way we describe here) tries to guess each of his upcoming moves, placing bets on the probability that he will sort one way or the other. In an attempt to optimize her success rate, she successively refines her predictions over the course of the game based on how each of her bets pays off.

Suppose the girl's brother begins the DCCS with a color sorting game. Initially, she might hedge her bets because she will not know whether the color cue (sort by red) and the shape cue (sort by star) predict the correct sort. However, because *red* always succeeds in predicting the correct sort in the color game, whereas *star* never does, she will place more and more of her chips on *red*, until she is no longer betting on *star* at all.

By the end of the first color game, she will have learned that she can maximize her returns by betting all her chips on *red* all the time. However, when the game changes—and her brother begins to sort by shape—her luck will change. She will still be placing her bets on *red*, but now *star* is proving to be the more reliable cue. In response to this, she will begin to divest some chips from *red* and put them on *star* instead. As the shape game continues, she will move more and more of her assets to *star*.

As we have described it so far, the process of reallocating chips will simply reverse itself on every switch trial: the girl will learn to bet on one cue, then another, and then back to the first. However, she can learn to predict her brother's sorts more accurately if she incorporates some contextual information into her strategy. In the DCCS, each sort takes place within the context of the shape or color game. Instead of placing her bets on only the two cues, she can hedge her bets by considering other possible cues. For example, in addition to the two-dimensional cues (*red* and *star*), she might consider placing her bets on conjunctive cues (Gluck & Bower, 1988) that integrate these dimensions with the contexts provided by the games (e.g., *color game* + *red* and *shape game* + *star*).

Because these conjunctive cues will only be present in the contexts in which they actually occur, they will not lead to losses: Each bet on the appropriate conjunctive cues will be rewarded. By contrast, the dimensional cues will only pay dividends half the time (*red* is a good bet in color games but not shape games, and vice versa for *star*). As evidence of the unreliability of the dimensional cues grows, the girl will move more and more of her chips away from them, as experience teaches her to focus her attention on the conjunctive cues instead.

A Model of Error-Driven Learning

To formally illustrate how this process leads to the pattern of perseveration seen in the DCCS, we simulated the competition between conjunctive (Contextual + Dimensional) cues representing *color game* + *red* and *shape game* + *star* and the individual cues *red* and *star* across repeated DCCS trials using the Rescorla and Wagner (1972) model of error-driven learning (Siegel & Allan, 1996; for other applications to learning in early childhood, see Ramscar et al., 2010; Ramscar, Dye, Popick, & O'Donnell-McCarthy, 2011).

In Rescorla–Wagner, learning is simulated by changes to the associative strengths between individual cues (C) and an outcome based on the results of discrete learning trials. If the presence of a cue or outcome X at time t is defined as present (X, t), and its absence as absent (X, t), then the predictive value V of a cue C_i for an outcome O after a learning event at time $t + 1$ can be stated as:

$$V_i^{t+1} = V_i^t + \Delta V_i^t, \tag{1}$$

while the change (Δ) to the predictive value of V_i^t can be defined as:

$$\Delta V_i^t = \begin{cases} 0 & \text{if ABSENT}(C_i,t) \\ \alpha_i \beta_1 \left(\lambda - \sum_{\text{PRESENT}(C_j,t)} V_j \right) & \text{if PRESENT}(C_j,t) \ \& \ \text{PRESENT}(O,t) \\ \alpha_i \beta_2 \left(0 - \sum_{\text{PRESENT}(C_j,t)} V_j \right) & \text{if PRESENT}(C_j,t) \ \& \ \text{ABSENT}(O,t) \end{cases} \tag{2}$$

Learning is thus governed by a discrepancy function in which λ is the total value of a predicted event O (the maximum amount of associative strength that the event can support; here it is simply set to 100, indicating that an event can be fully anticipated) and V_j is the predictive value for O given the set of cues C_j present at time t.

Importantly, there is no notion of explicit feedback in the model. Rescorla–Wagner learns—that is, adjusts cue weights—in response to the occurrence or *non*occurrence of relevant events. In this, it samples the implicit predictive structure of the learning environment. The learning rule is thus particularly well suited to capture the patterns of learning of young children, who are not yet able to selectively attend to their environment in the same manner as adults (Thompson-Schill et al., 2009).

In trials in which there is positive evidence—that is, in which expected responses *do* occur—the learning rule produces a negatively accelerated learning curve (the result of responses being better predicted, which reduces the discrepancy between what is expected and what is observed) and asymptotic learning over repeated trials (as responses become fully predicted).

In trials that result in negative evidence—that is, in which an expected response fails to occur—λ_j (the expected response) takes a value of zero because it did not occur. In such cases, the discrepancy function ($\lambda_j - V_j$) produces a negative value, resulting in a reduction in the associative strength between the cues present on that trial and the absent response j.

The total amount of predictive (cue) value any given outcome can support in learning is finite. Informally, this captures the idea that if predictive confidence keeps rising, it must eventually reach a point of certainty. As a result, cues *compete* with one another for relevance, and this produces learning patterns that often differ greatly from those that would arise by simply recording the correlations between cues and responses (Rescorla, 1988).

The rate of change (Δ) at t is determined by two factors: the overall learning rate β (where $0 \leq \beta \leq 1$) and the individual saliency of cues, α_i (where $0 \leq \alpha \leq 1$). In the simulation we conducted, $\lambda = 100\%$ (when j was present in a trial) or 0% (when j was not present in a trial), $\alpha_i = 1$ and $\beta_j = 0.3$. By setting the α_i parameter to 1, we effectively eliminated its influence on our simulations. These parameter settings were held constant throughout all the simulations reported below.

Simulating Competitive Learning in the DCCS

Simulation 1 models competitive learning in the DCCS over three successive games. The simulation predicts that if a child observes the correct sort being made on every trial, she will learn to value conjunctive (context + dimension) cues, allowing her to successfully switch rules in the DCCS. The initial implementation makes three straightforward assumptions:

1. The appropriate sorting response is always made; that is, the child learns in response to what *ought* to happen.
2. That the dimensional cues will each predict the correct sort on half the trials; for example, *red* will predict the correct sort in the color game but not the shape game.
3. That the conjunctive cues will always predict the correct sort; for example, *color game + red* will predict the correct sort in color game trials and will not be present on any trial in which shape is matched (i.e., shape game trials).

The model thus captures the state of a child who is given the opportunity to sort the cards, and either (a) correctly sorts each one on each trial, despite other possible responses, or (b) observes another participant make the correct response on every trial. Given this assumption, it is important to keep in mind the (somewhat counterintuitive) idea that even when the model predicts on one trial that the child will give an incorrect response, learning on the next trial proceeds *as if she had not*. The simulation thus makes predictions about what the child's response behavior will look like at any given point along the learning trajectory, after having observed—or made—a certain number of correct sorts and switches. (This assumption is discussed in further detail in the following section.)

Training began with a color game (Trial 1), followed by a shape game (appearing in Trial 7). To reflect children's experience in DCCS tests, the model was trained for six trials in each game, with training alternating between games.

In the first DCCS game shown in Figure 2, the dimensional cue *red* and the conjunctive cue *color game + red* gain in associative value. At the rule switch, the model pegs *red* as the favored cue to

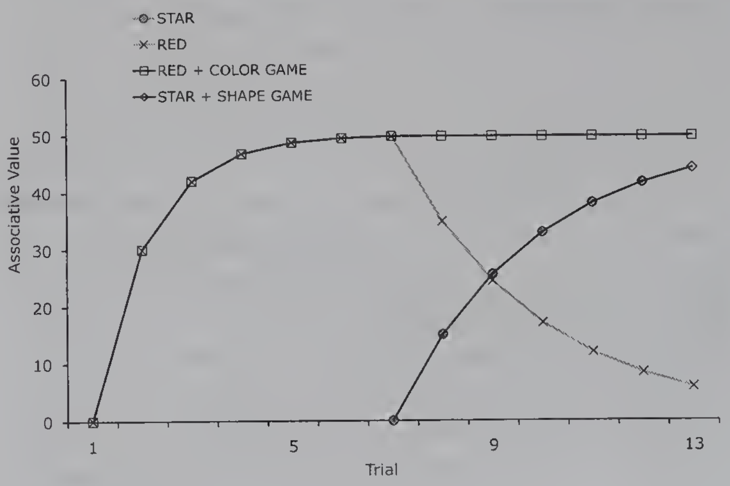


Figure 2. Rescorla–Wagner simulation of cue competition in two Dimensional Change Card Sort (DCCS) trials. Each line represents the association between a given cue and the correct response. Trials 1–6 represent a color game and Trials 7–12 a shape game (the six trials in each DCCS game correspond to six trials in the simulation). The value of *red* increases during the color game trials, meaning that in the first trial of the shape game, *red* is more highly valued as a sorting cue than either *star* or the conjunctive cue *star + shape-game*, leading to perseverative responding.

sorting. However, following the switch, *red* begins to lose associative value as it now fails to predict the correct sort. Over time, cue competition causes *red* to be effectively dissociated from correct sorting in this situation. The error generated by *red* over shape trials (where *red* continues to predict a color sort) results in a gradual shift of associative value from the dimensional cue *red* to the conjunctive cue *color game + red*, which does not lose value in this way because it is not present on shape trials.

It is important to note here that even though the dimensional cue (*red*) and the conjunctive cue (*color game + red*) co-occur with correct sorts with exactly the same frequency—that is, on half of all trials—the dimensional cues also co-occur with incorrect sorts, whereas the conjunctive cues do not. This means that the distribution of error (the combined evidence) will favor the conjunctive cues over the long run. Thus, the child will learn to ignore the unreliable cues over switch trials, thereby diminishing any bias toward perseverative responding.

This pattern of learning is clearly evident in the simulation (Figure 3): Depending on the game being played, the predictive value of a given conjunctive cue will either flatline if it is not present (e.g., *red + color game* on shape game trials, Trials 7–12) or increase if it is (e.g., *star + shape game* on shape game trials, Trials 19–24). Meanwhile, the value of each of the isolated dimensional cues *red* and *star* will continue to rise and fall depending on whether they are being reinforced in the game being played (e.g., *star* will rise on shape game trials and fall on color game trials). However, the overall trend for the dimensional cues is downward, as they will bleed associative value to the conjunctive cues. Indeed, after the second switch trial (Trial 13), the value of the dimensional cues will fall below that of the contextual cues, never to recover.

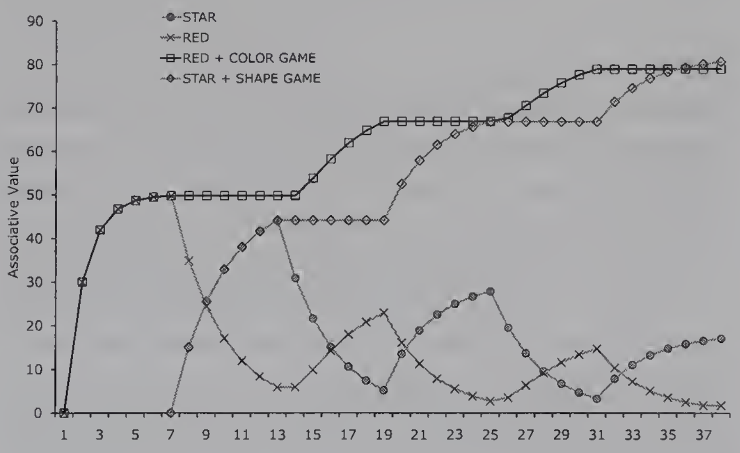


Figure 3. Rescorla–Wagner simulation of cue competition in six DCCS games. Each peak can be taken to represent a rule switch. The simulation assumes that the correct sorts are seen in each Trial. Prediction error (negative evidence) causes the individual cues to lose associative value to the conjunctive (context + dimension) cues over successive trials.

Behaviorally, the model suggests that perseverative sorting will occur whenever cues supporting the incorrect response are valued *over* those supporting the correct response, and, response conflict will occur whenever cues supporting the wrong response are strong relative to those supporting the correct response, even if they are not necessarily preferred. Given this, we can see that the model predicts a perseverative response on the first switch Trial (7): The value of the *red* cue far outstrips that of the *star* cue. However, the balance of power soon shifts. If we now turn to the third switch trial (19), we can see that while the value of the *red* cue outstrips that of the *star* cue, the value of the *star + shape game* cue far outstrips both dimensional cues.

Compatible Accounts

In Simulation 1, initial perseveration resulted from children failing to discriminate the contextual cues appropriate for correct sorting from competing cues that lead to error. This formal account of the way learning influences performance in the DCCS is compatible with a number of theoretical proposals about the factors that affect children's ability to successfully complete the task. For example, it has been suggested that perseverative responding in the task arises from attentional inertia: that once 3-year-olds have "focused their attention on a particular dimension, their attention gets stuck there, and they have extreme difficulty redirecting it" (Kirkham et al., 2003, p. 451). The problem is thought to be one of perspective: Once a child has grasped one sort dimension, she has trouble flipping her "mental focus" and thinking about it along the other (Diamond & Kirkham, 2005; Kloo & Perner, 2005). Our model of learning formalizes the idea of "mental focus," allowing the way attention shifts as a function of experience to be more precisely accounted for (see also Colunga & Smith, 2005).

Similarly, our account shares certain similarities with cognitive complexity and control theory (CCC; Zelazo et al., 2003), which suggests that successfully passing the DCCS requires that a child formulate a higher order "if . . . if . . . then" rule that integrates two conflicting pairs of rules within a propositional structure (e.g., "if a1, then if b1, then c1" while "if a2, then if b1, then c2"). Our model recharacterizes these rules in terms that allow us to examine the degree to which children's learning about them is implicit and contextual (Towse, Redbond, Houston-Price, & Cook, 2000). If children are given the DCCS without the aid of contextual training, they will still need to be able to resolve response conflict to pass it, as CCC theory suggests (indeed, most theories agree on this; see Diamond et al., 2005). However, formalizing the way that rules and representations are learned not only explains why children ordinarily fail the DCCS but also explains why children are able to respond flexibly in some tasks, even as they fail to do so in others. Furthermore, it allows us to predict the kind of learning experiences that should improve children's performance on the task.

Generalizing From Naming to Sorting

Simulation 1 (Figures 2 and 3) shows how contextual learning can serve to reduce response conflict in the standard DCCS, thereby offering an alternate route to success in the DCCS. If children are trained to associate sorting by shape with a "shape game" and sorting by color with a "color game," they can eliminate the response conflict normally associated with the DCCS by learning to value context-dependent cues to sorting (*red star card + color game = sort by red*) over context-independent cues, which frequently lead to errors.

This simulation provides a solution to a puzzle in the current literature: although under-fours routinely fail the DCCS, they can often reliably *name* the dimensions of the cards, even when they fail to correctly sort them (Kirkham et al., 2003). While on its face this finding may seem surprising, it need not be. Children have far more experience with naming than with sorting, and they get far more opportunities to watch other people use names in context. Given that children regularly hear objects referred to in terms of their shape and color, and so on, they are likely to have done a great deal of learning (including observational learning) about pairing shape words with shape questions and color words with color questions and so forth. (Landau, Smith, & Jones, 1988; Sandhofer & Smith, 1999). When it comes to a card-*naming* task, this kind of contextual discrimination learning will have diminished the degree to which color words suggest themselves as responses to shape-naming questions, thereby minimizing response conflict and enabling children to flexibly name the appropriate dimensions of the cards according to context.

Simulation 1 thus suggests that learning context-*dependent* cues could enable children to effectively bypass response conflict in the DCCS. Importantly, however, our computational simulation of *sorting* assumed that learners made the correct response throughout. Given that over the normal course of the DCCS, under-fours make frequent sorting errors, it is highly unlikely that a child playing the DCCS will actually experience the ideal learning conditions assumed in our simulation.

However, a considerable body of evidence supports the idea that learning generalizes across contexts (Gluck, 1991; Shepard, 1987). Within a discriminative learning framework, generalization results from a *lack* of discrimination. For example, in Simulation 1, when the initially poorly discriminated cues support perseverative responding, these perseverative responses are a form of (over-) generalization of learning from earlier trials (see Ramscar & Yarlett, 2007; Ramscar et al., 2010). Accordingly, if a child was to learn a particular configuration of cues and how they support (or failed to support) the correct response in one type of game context, what she learned about that context should transfer to other, similar contexts. This suggests that children's performance on the DCCS could be successfully enhanced by providing them with an opportunity to learn about a *color shape card + color game* context (or *color shape card + shape game* context) while playing a game at which they could succeed on every trial, even after a rule switch.

Because 3-year-olds can reliably and flexibly name the dimensions of the cards in the DCCS even when they fail to correctly sort by them, a naming game provides precisely this kind of learning opportunity. The results of Simulation 1 suggest that asking children to *name* the conflicting dimensions of the cards in the context of the games should lead to them learning the context-dependent cue configurations that lead to the correct naming responses. If children then generalize this knowledge to their *sorting* responses—using the same conjunctive cues they learned in the labeling game—this should substantially reduce the response conflict they encounter in the DCCS, markedly improving their chances for successful rule switching.

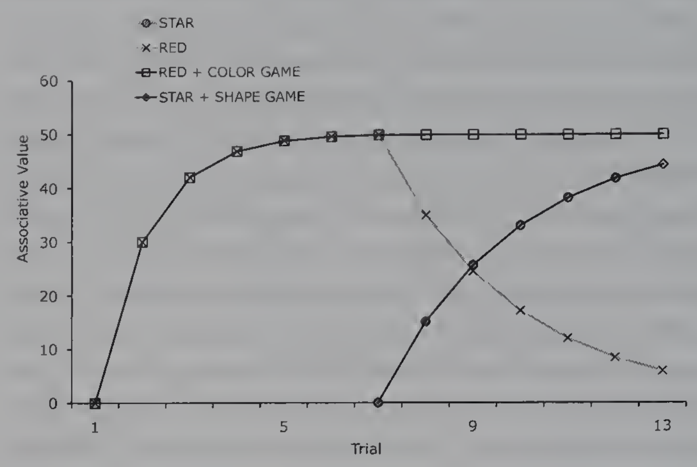
Simulating Noncompetitive Learning in the DCCS

Our simulation predicts that competitive discrimination learning about the game contexts in a labeling context will help children's performance on the DCCS. However, not all labeling tasks promote competitive learning. Depending upon the temporal sequence in which objects and their labels are encountered, learning can be made more or less competitive (Ramscar et al., 2010), thus allowing us to isolate the effects of competitive learning on DCCS performance.

In the naming game, presenting cards and then labeling them (feature-to-label, or FL-learning; Ramscar et al., 2010; Ramscar et al., 2011) allows the features of the cards and game contexts to serve as *cues* to labeling *responses*. This facilitates competitive learning, in which the various cues "compete" for relevance in predicting the appropriate labeling response, losing associativity to other, more reliable cues, and gaining associativity from other, less reliable cues. Cue competition thus allows a learner to discriminate the cues that most *reliably* predict the correct response (such as the conjunctive cues) from unreliable cues that result in error (such as the dimensional cues; see Figure 3).

Importantly, simply reversing the predictive (temporal) relation between labels and features can have a dramatic impact on cue competition. Whereas FL-learning involves predicting an individual label from a complex, multidimensional set of cues (the features of the cards), label-to-feature (or LF-) learning involves just the opposite: predicting a complex, multidimensional perceptual item (the card) from an individual cue (a symbolic label). Because verbal labels are sparse, they do not provide a learner with the raw materials required for discrimination learning (a matrix of feature dimensions that covary across learning contexts; Ramscar et al., 2010). In effect, LF-learning provides learners with just a single cue (a label) to learn from, thereby eliminating cue competition (if there are no other cues present, there are no other cues to lose associative value to). As a result, when a label acts as a cue, its associative value rises and falls according to whether it successfully or unsuccessfully predicts a given dimensional feature, effectively tracking the frequency with which labels and card features co-occur (for a formal analysis and extended discussion of the effects of information structure on learning, see Ramscar et al., 2010).

(a) FL-Simulation



(b) LF-Simulation

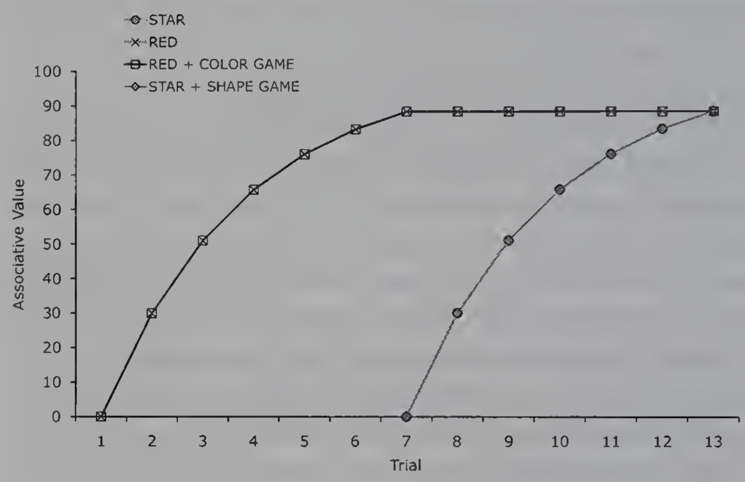


Figure 4. Simulations of cue competition in naming games in feature-to-label (FL) learning (left panel) and label-to-feature (LF) learning (right panel). Each line represents the association between a given cue and the correct response. Trials 1–6 represent a color-naming game, trials 7–12 a shape-naming game. FL-learning causes the *devaluing* of the individual dimensional cues (*red* and *star*), whereas LF-learning results in a uniform pattern of association between the dimensional and contextual cues.

To illustrate the effect this manipulation has on discrimination learning, Figure 4a depicts a variant of our earlier simulation in which cards depicting objects are alternately presented and then named in color and shape-naming games. As can be seen, FL-learning results in the devaluing of the individual dimensional cues (*red* and *star*), and learning of the conjunctive cues ("rules") appropriate to each naming game. However, in LF-learning (Figure 4b), because the cards' features do not compete as cues, the associative value accorded to the conjunctive cues will simply track the *frequency* with which dimensional features are present after labeling in each labeling game context.

As the dimensions occur with equal frequency (every card has both a color and shape), LF-learning will fail to discriminate between them. This means that although there will be learning about the game contexts in LF-presentations, LF-learning will provide identical evidence for the conflicting cues (Figure 4b). Thus, although a child asked to switch in a sorting task after LF-training will not face the same predicament as a child asked to switch dimensions in an ordinary DCCS task (where the available evidence actually *favors* perseveration; see Figure 2), she will still face the problem of selecting among competing responses.

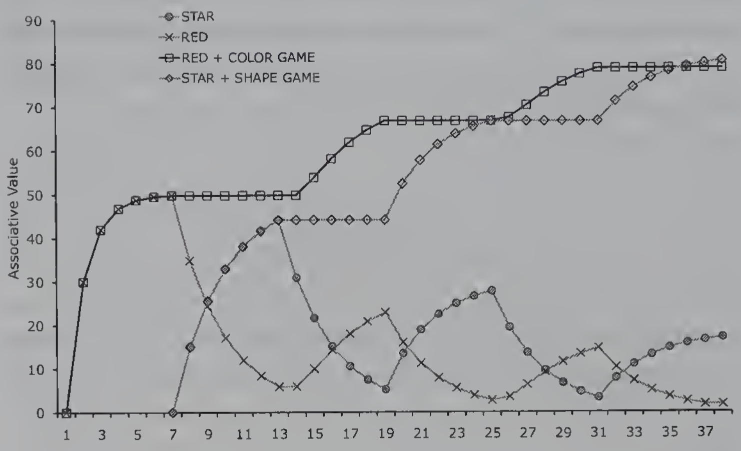
The different outcomes promoted by FL- and LF-training allow for our predictions about both the influence of discrimination learning in context, and the mechanisms that give rise to it, to be formalized and tested in detail. Figure 5 presents the results of two simulations that model performance on two DCCS tests. The simulations were given the same information—structured in the same way—as the children in our training experiment. In the first simulation, test performance was assessed following two FL naming games, and in the second simulation, following two LF naming games. In the games, two sets of cards that varied on two dimensions (shape and color) were presented and labeled in either an FL- or LF-sequence. In the color game, the color dimension was labeled, and in the shape game, shape was labeled.

Although the temporal sequence in which cards were presented and labeled varied across the two training conditions, the sequence of presentation and elicitation in DCCS testing was the same across conditions (children in all three experimental conditions were tested on the standard DCCS). Notably, because cards are presented prior to sorting in the DCCS, the information structure in test trials resembled the FL-training trials.

As can be seen in the FL-trained simulation (Figure 5a), by the first switch in sorting at Trial 19, the associative value of the appropriate conjunctive cue (*star + shape game*) is appreciably stronger than the competitor *red* cue, and by the second switch at Trial 31, the gap between reliable cue and competitor has more than doubled. This suggests that FL-training should do much to minimize response conflict, helping most children sort correctly across both tests.

By contrast, LF-training is far less helpful, at least initially (Figure 5b). At the critical first switch at Trial 19, the associative value of the conjunctive cue (*star + shape game*) is the same as that of the competitor *red* cue, which means that both of two conflicting responses are supported. However, by the second switch at Trial 31, there is a significant gap between the appropriate conjunctive cue (*red + color game*) and the competitor (*star*), which predicts reliable sorting. At first glance, this would seem to suggest that while many LF-learners should fail on the first switch, they may well pass on the second. Tempting as this interpretation is, it is also misleading, because the simulation assumed that *the correct sort is made throughout*. However, in the context of the experiment, this is only necessarily true of the training trials (1–12), in which children are labeling the cards. If, in testing, subjects are unable to resolve the response conflict on the first switch trial—and end up sorting incorrectly over Trials 19–24—they will gain none of the benefits from learning on subsequent trials that the simulation suggests.

(a) FL-Simulation



(b) LF-Simulation

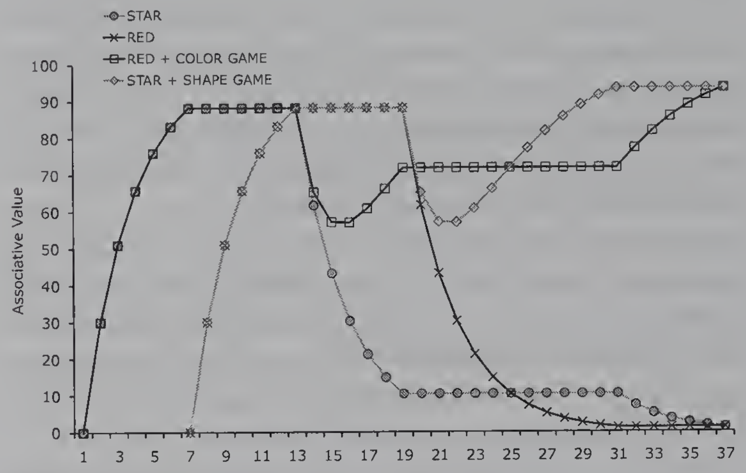


Figure 5. Simulations of Dimensional Change Card Sort (DCCS) testing following feature-to-label (FL) training (a), and label-to-feature (LF) training (b). In both simulations, Trials 1–12 depict training (naming games), while testing occurs over Trials 13–24 (first DCCS test) and 25–36 (second DCCS test). The important switch trials in testing occur on Trials 19 and 31. Breaking it down further, Trials 1–6 represent a color *naming* game, Trials 7–12 a shape-*naming* game, Trials 13–18 a color *sorting* game, and Trials 19–24 a shape *sorting* game. Because the initial sort dimension is reversed in the second DCCS, Trials 25–30 also represent a shape *sorting* game, with another switch back to color *sorting* in Trials 31–36. In interpreting these figures, and comparing them with our empirical results, it is important to note that the curves of the model represent the learned strength of a response, not the probability of a response. In situations in which there is response conflict, young children—unlike adults—tend to maximize, going with the most likely response every time, rather than probability match, alternating between responses based on their respective probabilities (Hudson-Kam & Newport, 2005).

Figure 5 thus confirms our prediction that FL-structured naming games ought to provide children with considerable help in avoiding perseverative responding (Figure 5a), whereas LF-training will only be of help *if* children can overcome the response conflict that arises on the critical first switch trial (Figure 5b).

Training Experiment

To test these ideas, we examined the effect that providing discrimination training to children would have on their performance in the DCCS.

Method

Participants. A total of 47 English-speaking children between 3 and 4 years old (M = 3 years 6.8 months) participated in this study, with a near-even balance between genders (25 girls and 22 boys). Participants were recruited from Stanford and the surrounding community.

Methods and materials. Two groups of children received either FL or LF training on the cards (Ramscar et al., 2010), before completing standard DCCS tasks (Zelazo, 2006). A control group was tested on the DCCS *without* training. All three groups were age and gender balanced.

In the two training conditions, children were introduced to shape and color games prior to the DCCS using 12 sorting cards (six yellow flowers and six green boats). In the FL-condition, children were first told, "In the shape game, we name the different shapes on these cards." The experimenter then presented the first card to the child and asked the child to label it. After children correctly labeled the first 6 of the 12 cards, the experimenter said, "We're going to play the color game. In the color game, we are going to say what colors are on these cards." Children then labeled the remaining six cards in the new game.

Whereas children in the FL condition saw the card and labeled it, children in the LF condition were asked to say the label *before* seeing the card. They were told, "In the shape game, we name the different shapes on these cards. The first card is going to be a flower—can you say 'flower'?" The experimenter showed the card to the child only after the child had repeated the label. The basic structure of the LF-training was the same as the FL-training: naming six cards by one dimension and then switching to the other dimension.

The two training groups and the control group then completed two standard DCCS tasks. The first DCCS task used the same sort cards as those in training (yellow flowers and green boats), whereas the second DCCS task used new cards, with different shapes and colors (blue trucks and red stars). The second DCCS task thus tested children's ability to generalize what they had learned, as it required them to sort by newly introduced colors in the context of the color game, rather than simply select yellow versus green.

The first testing dimension (shape or color) was counterbalanced across children, and children were required to correctly sort six cards in the preswitch. Once a child had done this, the sorting dimension was switched. Exactly six cards were sorted in the postswitch test. Before each trial, the subjects were either reminded of the current game's rules or asked to answer "knowledge questions," such as, "Where do the flowers go? Where do the boats go?" Children were given *no feedback* about their sorting of the cards. After the first DCCS task, children completed the second DCCS task with new cards, and the switching dimension was reversed from the first task.

Results

All children in the two training conditions correctly labeled the cards. In testing, children were considered to have "passed" the DCCS task if they sorted at least five of the six postswitch cards correctly. Sixty-nine percent of the FL-trained children passed the first DCCS task and 75% the second. By contrast, 33% of the LF-trained children passed the first task and 40% the second. Nineteen percent of the control children passed in each test (Figure 6).

Chi-square tests (the usual method of analysis for the DCCS) revealed significantly higher rates of successful rule switching in the FL-condition (11/16

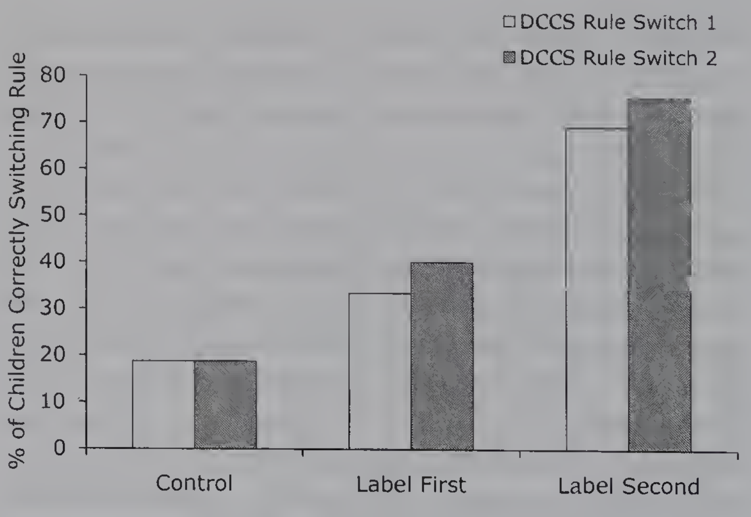


Figure 6. Percentage of children switching correctly in the first and second Dimensional Change Card Sort tasks by condition.

children succeeded) compared to 5/15 in the LF-condition, $\chi^2(1, N = 31) = 9.7$, $p = .005$; second switch, FL-condition (12/16 succeeded), LF-condition 6/15, $\chi^2(1, N = 31) = 17.0$, $p = .001$. Against the control group (3/16 in each), the comparisons with the FL-condition were, first switch, $\chi^2(1, N = 32) = 14.9$, $p = .001$; second switch, $\chi^2(1, N = 32) = 23.7$, $p = .001$. (These findings were confirmed in an equivalent series of parametric analyses.)

A series of one-sample *t* tests revealed that, overall, the FL-trained children had managed to sort correctly at above chance rates in the switch trials, $t(31) = 2.666$, $p < .02$, confirming that their training had enabled them to overcome the perseverative bias induced in the DCCS, while the baseline children sorted correctly at below chance rates, $t(31) = -3.161$, $p < .005$, confirming that they had perseverated on their prior sorts in the switch trials, just as previous studies of the DCCS in children of this age suggested they would.

Taken as a whole, the LF-trained children performed at chance in the switch trials, $t(29) = -0.586$, $p = .56$. This was consistent with the predictions of our LF-trained simulation, which suggested that while these children would learn something about the contextual rules, their use of the rules would be subject to considerable interference from the irrelevant cues (Figure 5b). However, inspection of the LF-trained children's results revealed that individually, the children's performance was not at chance; rather, some children were consistently sorting correctly, whereas others were perseverating.

Given that our maturational account suggests that response-conflict processing is acquired gradually, it follows that we should expect that, on average, LF-training ought to be of more benefit to the older children in our sample (older 3-year-olds are more likely to be in a position to process the limited conflict that remains after training than younger 3-year-olds). This should be especially true in comparison to the baseline children, who will still face a perseverative bias after the first sort game. In contrast, FL-learners ought to be *less* affected by age, as FL-learning leads to more distinct representations of the sort rules than LF (Figure 5).

To examine the extent to which these predictions were borne out in our results, we performed a median split by age on the groups of children in the study (for a breakdown of these groups, see Table 1), subjecting children's sorting to analysis in a 3 (training condition) × 2 (older vs. younger) × 2 (first vs. second DCCS trial) analysis of variance. This revealed that while there was no effect of the first versus the second DCCS test on children's performance, $F(1, 94) = 0.120$, $p > .7$, both training, $F(2, 93) = 9.199$, $p < .001$, and age, $F(1, 94) = 14.458$, $p < .001$, did have a significant effect on sorting. Furthermore, consistent with the predictions of our model, there was a significant interaction between age and training in children's performance, $F(2, 93) = 3.229$, $p < .05$ (see Figure 7). As expected, while the discrimination learning brought about by FL-training helped children pass the DCCS across the full age spectrum tested in our experiment, the overall reduction in response conflict in LF-training was *only* of benefit to the older children tested.

Notably, there may be other possible explanations for why the LF-trained children failed to benefit from training in the same way as their FL-trained counterparts. In particular, the FL-structure of testing may have made it easier for participants trained FL to generalize their knowledge to the sorting task. Given the design of the standard DCCS, this possibility cannot be completely ruled out. However, it should be noted that in previous studies of number and color learning in children, and relative pitch and complex-category learning in adults, FL-trained participants consistently outperform LF-trained participants, even when testing is structured LF (Ramscar et al., 2010; Ramscar et al., 2011). The consistency of these temporal ordering effects across the visual and auditory domains, and

Table 1
Mean Age and Population Size of the Groups in the Analysis of Age on Training and Performance

		FL	LF	Control
Younger	Mean age	3 years 2.5 months	3 years 2.5 months	3 years 2.3 months
	N	8	8	8
Older	Mean age	3 years 8.2 months	3 years 8.1 months	3 years 8.2 months
	N	8	7	8

Note. FL = feature-to-label; LF = label-to-feature.

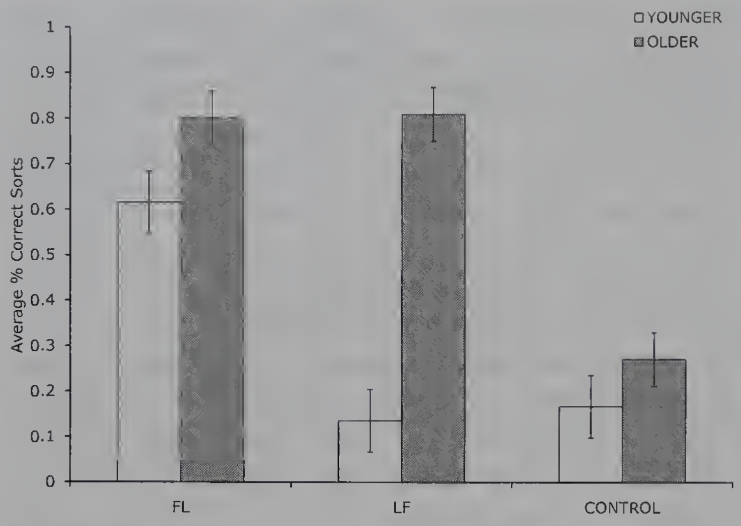


Figure 7. Percentage of correct sorts by the older and younger children by training condition. Error bars are SEM.

across a diverse array of tasks, argues for the consistency of the underlying learning mechanism.

Discussion

We predicted that playing naming games that promoted discrimination learning about the sorting rules used in the DCCS would improve children's performance on this task. Consistent with these predictions, younger children given training that was designed to promote contextual discrimination learning were consistently able to switch rules and pass the DCCS. When a less effective training program was administered, only the older children derived any benefit; children given the task without any prior training failed as expected. Taken together, these findings are consistent with our suggestion that the mind has multiple ways of dealing with conflicting response demands, and with evidence that online processing of response conflict appears to develop over the course of childhood (Rueda et al., 2005) and may thus be largely unavailable to children under four (Ramscar & Gitcho, 2007; Thompson-Schill et al., 2009). That under-fours are often able to behave in ways that are flexible and contextually appropriate appears to be due to discrimination learning. As the results of the FL-trained children show, under-fours can match their behavior to context in remarkably subtle and sensitive ways, but only once discrimination learning has enabled them to do so (for comparable effects of training in "theory of mind" tasks, see Amsterlaw & Wellman, 2006; Slaughter & Gopnik, 1996).

Applying a Dual Route Account to Other Findings

Our results and analysis suggest that two factors influence the way under-fours deal with potentially conflicting task demands, such as those imposed by the DCCS, namely, in terms of discrimination learning on one hand, and response-conflict processing on the other. There are at least half a dozen theoretical accounts of the effect that response conflict has on DCCS performance. However, given under-4s' overall weakness when it comes to response-conflict processing, it seems clear that an increased understanding of the way that children learn as they perform the task, and the contexts in which this learning occurs, can offer insight into a range of theoretical debates in the literature.

For example, in recent years, there has been much interest in "negative priming" (NP) effects in DCCS tasks. In a classic NP study, Zelazo et al. (2003) ran two variants of the DCCS: a partial change version and a total change version. In the partial change task, the dimension that was irrelevant to sorting in the preswitch was changed in the postswitch (e.g., children might sort *red* rabbits by shape during the preswitch and *yellow* rabbits by color during the postswitch). In the total change task, by contrast, both dimensions changed from preswitch to postswitch (e.g., children might sort *red rabbits* by shape during the preswitch and *blue trucks* by color in the post-switch). Zelazo et al. found that while most 3-year-olds continued to perseverate in the partial change task, most passed in the total change task. Zelazo and colleagues attributed these results to negative priming—that is, to the "disruption or slowing of a response to a stimulus that has previously been ignored" (Müller et al., 2006)—and concluded that to successfully sort during the preswitch trials, children needed to "suppress" attention to the irrelevant dimension, a tactic that then proved troublesome when that dimension become relevant again in the postswitch.

This theoretical stance is both in keeping with our suggestion that perseverative responding in the postswitch is the result of how dimensional cues are devalued in the preswitch and, too, with our suggestion that 3-year-olds generally lack the ability to process response conflict online. However, framing the different "values" of children's representations in terms of formal learning models, allows us to give a more nuanced explanatory account of both how performance develops over learning trials and how performance will be affected by various changes in the task paradigm.

Take the partial change version of the DCCS: In the preswitch shape game, a child learns that cards are sorted by *rabbit* but not by *red*; then, in the postswitch, a child is expected to sort *yellow rabbits* by color. In this version of the task, the child learns on the preswitch that the dimensional cue *rabbit* is a strong predictor of a correct sort. Then on the postswitch, rabbit competes with a new cue—*yellow*—which has no predictive value whatsoever, as it has not been learned about; indeed, if anything has been learned about *yellow*, it is that it is not an informative cue to sorting. That children perseverate, and continue to sort by *rabbits*, is readily captured by a learning account.

The results of the total change version are similarly predictable: In the preswitch, the child learns about a specific set of cues and their outcomes (red rabbits sorted by shape); whereas, in the postswitch, the child learns about an entirely new set of cues and their outcomes (blue trucks sorted by color). While we would expect some form of cue generalization even after one switch, and therefore some learning about color—which may explain why fully one fourth of 3-year-olds still do not pass this version—we would also expect that there should be much *less* interference when the sets of cues change completely, and correspondingly higher overall pass rates.

Separating out the influences of learning and response conflict can reconcile a number of other notable findings regarding the various factors that improve 3-year-olds' performance on DCCS tasks. For example, having the child—rather than the experimenter—label the card to be sorted (Towse et al., 2000) facilitates discrimination learning, whereas having the bins marked by cartoon characters (who desire certain cards) instead of target cards (with conflicting dimensions; Perner & Lang, 2002) reduces the response conflict inherent in the task. Likewise, separating out the relevant dimensions (color and shape) and having the child characterize them in relation to different objects (Diamond et al., 2005; Kloo & Perner, 2005) does both: reducing the task demands, while facilitating learning.

Considering both of these issues in tandem may help explain why even 4-year-olds encounter difficulty with flexible switching when strong situational cues are absent (Blaye, Paour, Bernard-Peyron, & Bonthoux, 2006), while even some of the youngest 3-year-olds in our experiment were able to pass the DCCS, given FL-training. In addition, the marked split between the older and younger children in our LF condition is a telling illustration of how cognitive maturity and learning interact in performance. These findings underscore the idea that while changes in response conflict processing around age 4 noticeably begin to facilitate flexible switching, the process of change is gradual, and subject to an extended developmental timetable (Anderson, 2002; Davidson et al., 2006; Rueda et al., 2005).

The Advantages of Cognitive Immaturity

In all of this, one final question remains: Why is the ability to dynamically process response conflict comparatively weak in young children? What possible benefit might this confer? We suggest that delayed cognitive maturation may actually be advantageous to children when it comes to learning behavioral conventions (Ramscar & Gitcho, 2007; Thompson-Schill et al., 2009). Both linguistic and social knowledge are, in essence, conventional. For symbols (such as words) to have "symbolic value," their values must be both conventionalized and internalized, and it may be that inflexible learning is actually useful in this regard (Ramscar et al., 2010).

The question of *how* whole communities come to share symbolic conventions is rarely posed, let alone answered. We propose that if young learners are unable to actively select what they attend to during the course of learning, then they will tend to sample the environment in much the same way: Given similar input to learn from, they will acquire similar representations of that input. If, on the other hand, those same learners were able to choose what they attended to in learning, the potential for conventions to be acquired in this way would decrease dramatically. It may be that the less children are able to direct their attention in learning, the more what they learn will be shaped by the statistical regularities of their physical, social, and linguistic environments (see also Hudson-Kam & Newport, 2005; Singleton & Newport, 2004). It follows, then, that poor performance on the DCCS may be part of the price children pay for their extraordinary abilities when it comes to acquiring social and linguistic conventions.

Finally, the ability of under-fours to "hide" their cognitive inflexibility, such that it requires careful experimentation to be revealed, and the speed with which children in our experiment were able to utilize contextual cues to succeed at the DCCS, may have important implications for our understanding of everyday cognitive processing in adults. The evidence presented here suggests that young children make extensive use of context in their interactions with the world. It may be that much of adult cognition works in the same way and is far more contextually embedded than is often supposed.

References

Allman, J. M., Watson, K. K., Tetreault, N. A., & Hakeem, A. Y. (2005). Intuition and autism: A possible role for Von Economo neurons. *Trends in Cognitive Science, 9*, 367–373.

Amsterlaw, J., & Wellman, H. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development, 7*, 139–172.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*, 71–82.

Blaye, A., Paour, J. P., Bernard-Peyron, V., & Bonthoux, F. (2006). Categorical flexibility in children: Distinguishing response flexibility from conceptual flexibility: The protracted development of taxonomic representations. *European Journal of Developmental Psychology, 3*, 163–188.

Brooks, P. J., Hanauer, J. B., Padowska, B., & Rosman, H. (2003). The role of selective attention in preschoolers' rule use in a novel dimensional card sort. *Cognitive Development, 18*, 195–215.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Cepeda, N. J., & Munakata, Y. (2007). Why do children perseverate when they seem to know better: Graded working memory, or directed inhibition? *Psychonomic Bulletin & Review, 14*, 1058–1065.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*, 347–382.

Colunga, E., & Smith, L. B. (2008). Flexibility and variability: Essential to human cognition and the study of human cognition. *New Ideas in Psychology, 26*, 174–192.

Davidson, M. C., Amso, D., Cruess Anderson, L., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 271–327). San Diego, CA: Academic Press.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). London, UK: Oxford University Press.

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the Dimensional Change Card Sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology, 28*, 689–729.

Diamond, A., & Kirkham, N. Z. (2005). Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychological Science, 16*, 291–297.

Erickson, K. I., Milham, M. P., Colcombe, S. J., Kramer, A. F., Banich, M. T., Webb, A., et al. (2004). Behavioral conflict, anterior cingulate cortex, and experiment duration: Implications of diverging data. *Human Brain Mapping, 21*, 98–107.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science, 2*, 50–55.

Gluck, M., & Bower, G. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27*, 166–195.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching. *Developmental Science, 13*, 622–635.

Hudson-Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology, 387*, 167–178.

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental Psychology, 35*, 770–780.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behaviour*. New York: Appleton-Century-Crofts.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449-467.

Kloo, D., & Perner, J. (2005). Disentangling dimensions in the Dimensional Change Card Sorting task. *Developmental Science, 8*, 44–56.

Kloo, D., Perner, J., Kerschhuber, A., Dabernig, S., & Aichhorn, M. (2008). Sorting between dimensions: Conditions of cognitive flexibility in preschoolers. *Journal of Experimental Child Psychology, 100*, 115–134.

Landau, B., Smith, L. B., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*, 299–321.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature, 431*, 760–767.

Moriguchi, Y., & Hiraki, K. (2009). Neural origin of cognitive shifting in young children. *Proceedings of the National Academy of Sciences, 106*, 6017–6021.

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage in childhood. *Developmental Psychobiology, 40*, 255–265.

Müller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in the Dimensional Change Card Sort task. *Child Development, 77*, 395–412.

Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*, 335–337.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass, 4*, 906–924.

Perner, J., & Lang, B. (2002). What causes 3-year olds' difficulty on the Dimensional Change Card Sorting task? *Infant & Child Development, 11*, 93–105.

Perner, J., Stummer, S., & Lang, B. (1999). Executive functions and theory of mind: Cognitive complexity or functional dependence? In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 133–152). Mahwah, NJ: Erlbaum.

Ramscar, M., Dye, M., Popick, H., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE*, e22501.

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends In Cognitive Science, 11*, 274–279.

Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science, 31*, 927–960.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). Feature-label-order effects and their implications for symbolic learning. *Cognitive Science, 34*, 909–957.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist, 43*, 151–160.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccamanno, L., & Posner, M. I. (2005). Training, maturation and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences, 102*, 14931–14936.

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35*, 668–679.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology, 57*, 87–115.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317–1323.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*, 207–218.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin and Review, 3*, 314–321.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American sign language from inconsistent input. *Cognitive Psychology, 49*, 370–407.

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*, 2967–2988.

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually-cued attention and early word learning. *Cognitive Science, 34*, 1287–1314.

Thompson-Schill, S. L., Ramscar, M., & Evangelia, C. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science, 8*, 259–263.

Towse, J. N., Redbond, J., Houston-Price, C. M. T., & Cook, S. (2000). Understanding the Dimensional Change Card Sort. Perspectives from task success and failure. *Cognitive Development, 15*, 347–365.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412*, 43–48.

Yerys, B. E., & Munakata, Y. (2006). When labels hurt but novelty helps: Children's perseveration and flexibility in a card-sorting task. *Child Development, 77*, 1589–1607.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931–959.

Yeung, N., Nystrom, L. E., Aronson, J. A., & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *Journal of Neuroscience, 26*, 1429–1438.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function. *Monographs of the Society for Research in Child Development, 68*(Serial No. 274).

Child Development, July/August 2013, Volume 84, Number 4, Pages 1324–1337

Increases in Cognitive and Linguistic Processing Primarily Account for Increases in Speaking Rate With Age

Ignatius S. B. Nip
San Diego State University

Jordan R. Green
University of Nebraska–Lincoln

Age-related increases of speaking rate are not fully understood, but have been attributed to gains in biologic factors and learned skills that support speech production. This study investigated developmental changes in speaking rate and articulatory kinematics of participants aged 4 ($N = 7$), 7 ($N = 10$), 10 ($N = 9$), 13 ($N = 7$), 16 ($N = 9$) years, and young adults ($N = 11$) in speaking tasks varying in task demands. Speaking rate increased with age, with decreases in pauses and articulator displacements but not increases in articulator movement speed. Movement speed did not appear to constrain the speaking. Rather, age-related increases in speaking rate are due to gains in cognitive and linguistic processing and speech motor control.

The rate at which we speak provides information to others about our gender, dialect, cognition, health, and age. For instance, young children and even adolescents talk at a slower rate than adults (Amster, 1984; Haselager, Slis, & Rietveld, 1991; Kowal, O'Connell, & Sabin, 1975; Sturm & Seery, 2007; Walker, Archibald, Cherniak, & Fish, 1992). The prolonged development of speaking rate is particularly surprising given that adolescents have already established the rules and sounds of their language (Kowal et al., 1975). The factors that allow for increased rates of speech with age are not fully understood, but have been attributed to gains in biologic factors (e.g., anatomic growth, neurologic and neuromuscular maturation) and learned skills that support rapid spoken language production (motor learning; semantic, lexical, and phonologic access; and motor programming and planning). At the level of speech motor control, talkers have only three options to increase their rate of speech: (a) decrease the extent of oral movement, (b) increase the speed of oral movement, or (c) decrease the number and duration of pauses (Campbell & Dollaghan, 1995). Investigating how these variables and speaking rate covary with age will have implications for understanding the contributions of biologic and spoken language processing factors on speaking rate development.

Evidence for Constraining Biologic Factors

One untested biologic hypothesis is that the slowed rate of speech in children is, in part, due to relatively slow central and peripheral neural conduction speeds. Findings from studies on limbs using transcranial magnetic stimulation suggest that conduction times in children are limited by the fastest corticomotoneuronal efferent nerves (Müller & Hömberg, 1992) with central conduction times decreasing significantly with age (Müller, Hömberg, & Lenard, 1991). The conduction velocities of neural pathways innervating orofacial structures may also increase with age. Indirect evidence comes from a small study that demonstrates that the latency time of the perioral reflex of adults are faster than those of children (Barlow, Finan, Bradford, & Andreatta, 1993). Support for a limiting role of movement speed on the development of speaking rate would be provided by the observation of a strong association (i.e., a correlation of greater than 0.8, as defined by Cohen, 1988) between age-related changes in movement speed and speaking rate. More specifically, if speed is observed to be slow in children across both very simple and complex

---

This research was supported in part by grants from the National Institute on Deafness and Other Communication Disorders (NIH-NIDCD) R01 DC00646 and R01 DC009890, the Barkley Trust and the Nebraska Speech-Language-Hearing Foundation. The authors would like to thank Nirmal Srinivasan, Cynthia Didion, Lauren Brockevelt, Laura Beshaler, Antje Mefferd, Paige Leising, Wendy Quach, Susan Fager, Lacey LaBarge, Cory Quinlan, Cara Ullman, and Julie Wyss for their help in data collection and analysis.

Correspondence concerning this article should be addressed to Ignatius Nip, School of Speech, Language, and Hearing Sciences, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-1518. Electronic mail may be sent to inip@mail.sdsu.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0016
DOI: 10.1111/cdev.12052

speaking tasks, this finding would suggest that the changes in movement speed are predominantly biologically rather than linguistically driven.

Evidence of Motor Learning

Immature motor control is characterized by inefficiencies in motor control (Goldfield, Kay, & Warren, 1993). For example, in early reaching, poor motor control results in an overshoot of the displacement of the arm and hand (Jeannerod, 1988) and poor force regulation results in excessive pincer force during grasping (Potter, Kent, Lindstrom, & Lazarus, 2006). Poor position control in early speech development may similarly result in articulatory target overshoot, specifically, excessively large articulatory displacements (Green & Nip, 2010). Immature control of the speech motor system has been observed in young children, who produce larger and slower articulator movements than adults do (Goffman & Smith, 1999; Green & Wilson, 2006; Riely & Smith, 2003; Smith & Gartenberg, 1984; Smith & Goffman, 1998; Smith & Zelaznik, 2004; Walsh & Smith, 2002). The relatively large oral excursions produced by children may account for the slower rate of speech, particularly if they are not accompanied by proportional increases in movement speed. This finding would provide support for the suggestion that an essential process in the development of speaking rate is the optimal tuning of the speech motor control system through motor learning.

Evidence for a Reduced Capacity to Formulate Spoken Language

Speech motor control, including speed of articulator movement and consequently speech, is also dependent on cognitive, linguistic, and motor workloads (Green & Nip, 2010). The processing demands on spoken language production that are imposed by different speaking tasks may vary depending on factors including attention, utterance familiarity (word frequency and phonotactic probability), utterance length, and syntactic complexity. For example, the slowing of speech during the performance of a manual visuomotor tracking task is evidence of the influence of attentional demands on speech motor control (Dromey & Benson, 2003). In addition, children speak faster during simple speaking tasks, such as the repetitions of simple syllables, than during more demanding speaking tasks, such as conversational speech (Haselager et al., 1991). The relations between speaking rate and task demands suggest that children speak slower than adults, in part, because their articulator movement speeds are slowed by their reduced capacity to formulate spoken language. Therefore, the influence of speaking task demands on children's rate of speech would be supported by the observation that children's speeds of articulator movement are faster for low-demand speaking tasks, but slower for high-demand speaking tasks across development.

Another way to determine the contribution of limited spoken language processing on the development of speaking rate is to examine age-related changes in pausing patterns. Prior research has shown that children gradually reduce the number and length of pauses, a change that has been attributed to developmental gains in cognitive and linguistic processing (Kowal et al., 1975). A talker pauses more frequently and for longer as speaking tasks become more cognitively and linguistically demanding (Greene, 1984; Greene & Cappella, 1986). For example, adults pause longer, thereby decreasing their speaking rate, while speaking on a topic spontaneously than while speaking on the same topic after preparing an outline (Mitchell, Hoit, & Watson, 1996); similarly, young children speak more slowly in spontaneous speech as compared to familiar utterances (Walker & Archibald, 2006). The observation that changes in pausing, rather than in articulator movements, primarily account for age-related changes in speaking rate would suggest that gains in spoken language processing play a more prominent role in the development of speaking rate than do biologic factors.

Research Questions

The current study investigated age-related changes in speaking rate, articulatory kinematics (i.e., displacements and speeds), and pause duration across several speaking tasks that varied from simple alternating oral movements to story retell. The stimuli were designed to elicit different levels of processing demands (motor, linguistic, or cognitive processing) on spoken language production. Our central research question was to determine whether changes in speaking rate are primarily due to the constraining effects of biologic factors (movement speed), motor learning (refinement of displacements) or due to gains in spoken language processing (interactions between age, task demands, speaking rate, and pausing patterns). Evidence for biologic constraints on the development of speaking rate would be supported by two possible findings: (a) the observations that the speed of speech movements are not significantly different across all ages and tasks (which vary in processing demands), or (b) age-related changes in movement speed are not significantly different across tasks. In either scenario, the common course of movement changes cannot be attributed to differences across tasks, but instead to underlying biologic factors. Evidence for motor learning factors would be supported if oral excursions decrease with age and do not change with task demands. In contrast, observations of task-specific changes in movement speed (i.e., speed varies predictably across tasks) would provide evidence that spoken language processing (i.e., motor, linguistic, and cognitive) may drive developmental changes in speaking rate. For example, we might anticipate that the movement speed for simple syllables will change little with age, whereas movement speed would increase significantly for more challenging speaking tasks. The observation that articulatory displacements become smaller as speaking rate increases with age would provide support for the suggestion that an essential process in the development of speaking rate is the optimal tuning of the speech motor control system through motor learning. Finally, a complementary analysis was performed on connected speech samples at each age to further discern the relative contributions of biologic, motor learning, and spoken language production processing changes on the development of speaking rate. The suggestion that speaking rate increases are primarily due to increases in spoken language production processing would be supported by observations of a significantly greater change in pause duration than articulatory kinematics or a disassociation between age-related changes in speaking rate and articulatory kinematics.

Method

Participants

Data were collected on a total of 78 participants; however, 24 participants were not included in the analysis because they were either unable to complete the tasks, were unwilling to have the markers placed on their faces, or were identified as having a speech or language delay during the speech and language screening. As a result, 54 participants in the following age groups are included in this report: 4- (5 males, 2 females, *M* = 4.45 years, *SD* = 0.39 years), 7- (5 males, 5 females, *M* = 7.57, *SD* = 0.29 years), 10- (5 females, 4 males, *M* = 10.40 years, *SD* = 0.27 years), 13- (3 females, 4 males, *M* = 13.31 years, *SD* = 0.27 years), and 16-year-olds (5 females, 4 males, *M* = 16.39 years, *SD* = 0.32 years) and adults (5 males, 6 females, *M* = 23.75 years, *SD* = 3.52 years). All participants lived in Nebraska, spoke American English with a standard Midwest dialect as their first language, and were primarily from middle-class backgrounds. The majority of participants identified as White, not Hispanic. Four participants identified as White and Hispanic. Four participants did not identify their race as White. One identified as African American, not Hispanic; the second identified as Native American and White, not Hispanic; another identified as White, African American, and Hispanic, and the last identified as White, African American, Native American, and White, not Hispanic. None of the participants had any history of speech, language, hearing, or learning difficulties. Each participant passed a hearing screening at 1, 2, and 4 kHz at 20 dB HL for children and at 25 dB HL for adults, following American Speech-Language-Hearing Association (ASHA) guidelines for audiological screening (ASHA Panel on Audiologic Assessment, 1997). Child participants also passed a receptive and expressive language screening; 4-year-olds were screened using the Core Language subtests (Sentence Structure, Word Structure, Expressive Vocabulary subtests) of the Clinical Evaluation of Language Fundamentals–Preschool (CELF–P) 2nd edition (Wiig, Secord, & Semel, 2004). The language skills of the 7-, 10-, 13-, and 16-year-olds were screened using the Clinical Evaluation of Language Fundamentals–4 (CELF–4) Screening test (Semel, Wiig, & Secord, 2004). Based on an informal conversational speech sample, a certified speech-language pathologist did not identify and articulation delays or disorders.

Data Collection

Lip and jaw movements were captured using a three-dimensional motion capture system (Motion Analysis, Ltd., Santa Rosa, CA). This optically based system has eight infrared cameras (Eagle) with 1.3 megapixel resolution that were calibrated to the manufacturer's specifications prior to each data collection session. The system used a sampling rate of 120 frames per second for each camera. Fifteen spherical reflective markers, approximately 2 mm in diameter, were placed on the forehead, eyebrows, nose, lips and jaw with double-sided hypoallergenic adhesive tape as shown in Figure 1. A rigid plate with a miniature microphone and four markers was also attached to each subject's forehead. The forehead marker array was used for the off-line subtraction of head movements from those of the lower lip marker. High-resolution digital video and audio (44.1 kHz and 16 bits) signals of the participants were also recorded and used for parsing of bilabial opening and for acoustic analyses, as described below.

Participants were seated in front of the camera system. Video clips providing instructions and demonstrations of the speech tasks were projected on a large screen in front of the participant. The video clips ensured that the instructions (detailed below under "Speaking Tasks") and demonstrations were presented in an identical manner and uniform rate to each participant. For younger participants, an investigator also sat in front of the child in order to keep him or her facing the cameras. Occasional coaching on tasks was provided, particularly for younger children.

All movement traces were low-pass filtered at 7 Hz using a Butterworth filter (Butterworth, 1930). To obtain lip and jaw movements, the 3D Euclidean distance between the lower lip marker and the head marker at the right top head was calculated for each task. This calculation effectively subtracted head movement from the movement of the lower lip marker.

The initial opening gesture of the lower lip marker for a voiced bilabial segment was examined in order to compare the same movement gesture across all tasks. Specifically, the syllable of /b/ followed by an open vowel was parsed from the narrative retell ("**Bo**bby"), sentence repetition ("buy **Bo**bby"), syllable repetition ("uh**ba**"), and diadochokinetic or diadochokinetic (DDK) rates ("**buh**"). These syllables were then measured using custom algorithms developed for Matlab. Zero-crossings in velocity signals were used to detect the onset and offset of each opening gesture associated with each vowel. Displacement (mm) was also measured and was defined as the difference between the position of the lower lip marker at bilabial closure and at full opening for the vowel. Measures of peak speed (mm/s) and duration (s) were also obtained.

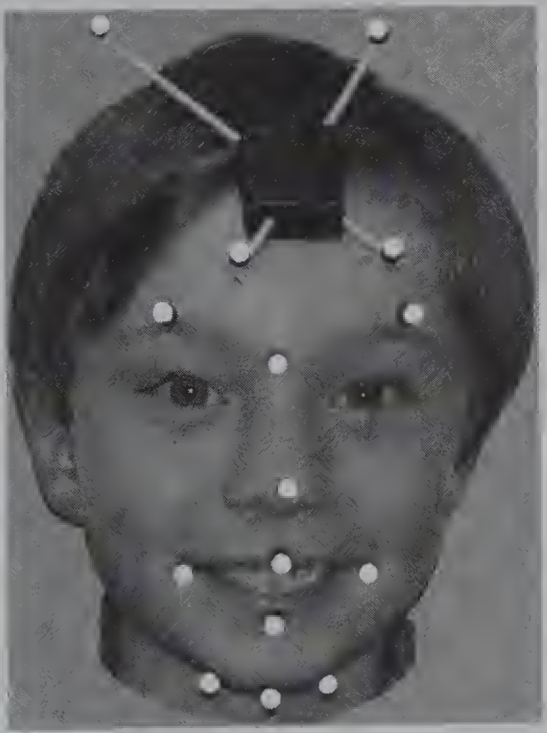


Figure 1. Marker set used for articulatory movement recording.

Speaking Tasks

All speaking tasks were purposefully loaded with bilabial segments to engage use of the lips and jaw. The large number of voiced segments allowed a Matlab-based algorithm (Green, Beukelman, & Ball, 2004) to automatically identify pauses in the speech signals. Four different speaking tasks were used to vary the amount of cognitive, linguistic, and motor processing demands. Speaking tasks included a narrative retell task using the "Bats, Boots, and Beets" story (Green, Nip, Wilson, Mefferd, & Yunusova, 2010), sentence repetition ("Buy Bobby a puppy"), syllable repetition ("uhba"), and a DDK task. For the narrative retell, participants were shown a picture corresponding to each sentence of the story as the story was read to them. They were then asked to retell the story while viewing the same pictures. During the DDK task, participants were asked to repeat "buh" fast and clearly and to do as many repetitions as possible in a single breath. For all tasks, speakers were asked to produce speech at their habitual rate and loudness.

The DDK, a speech-like task used by speech-language clinicians to examine the fastest alternating motion rates in the oromotor system, had the fewest task demands (Ziegler, 2002). Therefore, this task was used to identify the maximum repetition rate during a task that was relatively unencumbered by cognitive or linguistic processing demands or by motor planning (Haselager et al., 1991), providing a measure of the capacity of the participant's speech motor system (Ziegler, 2002). Relative to the DDK task, the syllable task was thought to increase linguistic processing demands because it imposed a syllabic structure (i.e., a vowel-consonant-vowel [VCV] sequence that conforms to English phonology rather than a continuous string of CV sequences) on the phonemic sequence. To further increase cognitive and linguistic demands, the simple sentence task was used to impose semantic and syntactic structure. Finally the narrative retell task was considered to be the most complex as it required participants to remember the story, in addition to producing a sequence of sentences in a simple narrative structure.

Analyses

All speech samples were orthographically transcribed and speaking rates were calculated for each task in syllables per minute. The rate in which speech is produced has been measured in several different ways in previous studies. In the fluency literature, speaking rate is commonly defined as the number of fluent and disfluent syllables divided by the amount of time to produce that utterance (e.g., Hall, Amir, & Yairi, 1999; Kelly & Conture, 1992). In contrast, articulation rate of the same utterance is the number of syllables divided by time during fluent speech (e.g., Kelly & Conture, 1992; Ryan, 2000). Studies in speaking rate of typically developing children have typically measured speaking rate as the number of syllables or words produced divided by the total time, including pausing, needed to produce that utterance (e.g., Flipsen, 2002, 2003). Articulation rate in these studies typically refer to the number of syllables or words produced divided by the duration, with pausing time removed (e.g., Flipsen, 2002, 2003; Walker & Archibald, 2006). As all the speakers in the current study were typically-developing individuals with no history of speech or language difficulties, speaking rate was operationally defined as the amount of time to produce an utterance, including pause times; articulation rate was operationally defined as the amount of time to produce an utterance with any pausing removed. The percentage of time for pausing measured in order to examine processing time a participant needed to generate an utterance.

Custom Matlab algorithms were used to determine the total pause time, defined as silence for 100 ms or longer but are not stop gaps (Green et al., 2004), for all speech tasks. For each audio file, an amplitude threshold (Green et al., 2004) was specified to allow the algorithm to determine the minimum acoustic energy needed to distinguish the speech signal from noise. Occasionally, in less than approximately 10 occasions, stop gaps (or silences) for the production of /b/ were identified as being pauses. These were discarded from the final analysis as it did not represent linguistic and cognitive processing. Previous studies (e.g., Haselager et al., 1991) have shown that decreases in pausing time may account for some of the developmental changes in speaking rate. No pausing was found for the DDK, syllable, and sentence task; only the narrative retell was found to contain interutterance pauses. This finding was anticipated and suggested that the DDK, syllable, and sentence task did not require either a respiratory or cognitive pause even in the 4-year-old children. Therefore, only the narrative retell task was analyzed for pausing time.

Statistical Analyses

For each participant, means of the peak speeds, displacement, and duration were measured across the multiple repetitions of each speaking task for the lower lip marker. The means for peak speed, displacement, and duration were Winsorized using the 25th and 75th percentile Tukey hinges to calculate the upper and lower bounds of the distribution for a speaking task for each age. Observations of peak speed, displacement, and duration that fell outside the calculated upper and lower bounds were replaced with the calculated maximal or minimal values.

Bivariate correlations between age, sex, speaking rate, peak speed, duration, and displacement for the DDK, syllable repetition, and sentence repetition tasks were conducted. Bivariate correlations using the same variables and percent pausing time were also conducted for the narrative retell task. Statistical analyses tested for the effects of age and speaking task (DDK, syllable, sentence, narrative) on speaking rate, peak speed, duration, and displacement. A multivariate analysis of variance (MANOVA) was used to determine if age and task cause significant changes in speaking rate, peak speed, displacement, and duration. Separate two-way univariate analyses of variance were then used determine significant differences in speaking rate, peak speed, duration, and displacement between age groups and tasks.

Results

Bivariate correlations between age, sex, speaking rate, peak speed, duration, and displacement for each of the speaking tasks are shown in Tables 1–4. Table 1 presents these correlations for the DDK task, Table 2 for the syllable repetition task, and Table 3 for the sentence repetition task. Percent pausing time, considered to be an indicator of cognitive, linguistic, and motor processing, was calculated only during the narrative task, as the other

Table 1
Correlations Between Age, Sex, Speaking Rate, Peak Speed, Duration, and Displacement for the Diadochokinetic Task

	Sex	Speaking rate	Peak speed	Duration	Displacement
Age	.09	.75***	.16	.69***	−.05
Sex		.11	−.07	−.11	−.13
Speaking rate			−.13	−.95***	−.39*
Peak speed				.11	.95***
Duration					.38*

*$p < .05$. ***$p < .001$.

Table 2
Correlations Between Age, Sex, Speaking Rate, Peak Speed, Duration, and Displacement for the Syllable Repetition Task

	Sex	Speaking rate	Peak speed	Duration	Displacement
Age	.11	−.01	.40*	−.29*	.20
Sex		.15	.06	.07	.05
Speaking rate			−.10	−.28*	−.15
Peak speed				.21	.95**
Duration					.46**

*$p < .05$. **$p < .01$.

tasks were all repetition tasks with no inherent interword pauses. Correlations between age, sex, speaking rate, peak speed, duration, displacement, and percent pausing time for the narrative retell task are shown in Table 4. Because sex was not found to be significantly correlated with any variable, it was removed from further analyses.

A MANOVA was conducted to examine the multivariate difference by age and task on the dependent variables speaking rate, articulation rate, peak speed, displacement, and duration. A significant Age × Task interaction was found, Wilks = .005, $F = 7661.61$, $p < .001$. Both the main effect of age, Wilks = .332, $F = 9.48$, $p < .001$, and task, Wilks = .102, $F = 43.67$, $p < .001$, were significant. Follow-up analyses of variance for each dependent variable are reported below.

Table 3
Correlations Between Age, Sex, Speaking Rate, Peak Speed, Duration, and Displacement for the Sentence Repetition Task

	Sex	Speaking rate	Peak speed	Duration	Displacement
Age	.11	.43**	.50**	−.32*	.34*
Sex		.01	.02	−.02	.03
Speaking rate			.16	−.51**	−.06
Peak speed				−.12	.90**
Duration					.28*

*$p < .05$. **$p < .01$.

Age and Task Effects on Speaking Rate and Articulation Rate

Speaking rate and articulation rate for each task and age is shown in Figure 2. A between-groups factorial analysis of variance with follow-up analyses using the least significant difference (LSD) minimum mean difference procedure was performed to examine the effects of age and task on speaking rate. A significant main effect of age $F(5, 188) = 20.81$, $p < .001$ was found, with speaking rate increasing with age up until 13 years of age. The main effect of task was also statistically significant $F(3, 188) = 108.04$, $p < .001$; however, these main effects were qualified by a significant interaction of task and age, $F(15, 188) = 3.20$, $p < .001$.

The LSD minimum mean difference was used to follow up the Age × Task interaction. The 4-, 10-, 13-year-olds demonstrated the same pattern; narratives had significantly slower speaking rates than the other tasks and the DDK task had significantly higher speaking rates than all the other tasks. For the 7-year-olds, narratives similarly had slower speaking rates than the other tasks. In addition sentences were also found to have significantly lower speaking rates than DDK and syllables. The 16-year-olds also demonstrated significantly higher speaking rates for DDK than the other tasks and sentences had significantly higher speaking rates than narratives. Adults were similar to 16-year-olds; they demonstrated significantly higher speaking rates for DDK than the other task but sentences had higher speaking rates than both syllables and narratives.

Next, in examining age differences within a task, For the DDK task, 4-year-olds were shown to significantly lower speaking rates than did the other age groups, 7-year-olds had significantly lower speaking rates than did 13-year-olds, and adults had significantly higher speaking rates for every age group except the 13-year-olds. For the syllable task, 7-year-olds had significantly higher speaking rates than did 4-, 10-, 16-year-olds, and adults; 13-year-olds had significantly higher speaking rates than did 4- and 10-year-olds. The sentence repetition task demonstrated that 13-year-olds had significantly higher speaking rates for 4-, 7-, and 10-year-olds; 16-year-olds had significantly higher speaking

Table 4
Correlations Between Age, Sex, Speaking Rate, Peak Speed, Duration, Displacement, and Percent Pausing Time for the Narrative Task

	Sex	Speaking rate	Peak speed	Duration	Displacement	Pct. pausing
Age	.11	.78**	.44**	−.75**	.11	−.65**
Sex		−.01	−.14	−.09	−.18	−.06
Speaking rate			.33*	−.67**	.04	−.67**
Peak speed				−.44*	.89**	−.37*
Duration					−.03	.43**
Displacement						−.19

*$p < .05$. **$p < .01$.

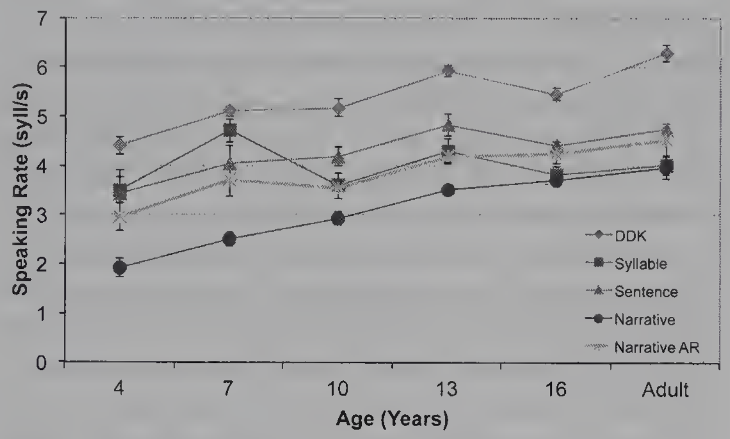


Figure 2. Speaking rate in syll/s for diadochokinetic (DDK), syllable repetition, sentence repetition, and narrative retell and articulation rate (AR) in syll/s for narrative retell. ARs for diadochokinetic, syllable repetition and sentence repetition tasks were the same as the speaking rates.

rates than 4-year-olds; and adults had significantly higher speaking rates than did 4- and 7-year-olds. Finally, the narrative task revealed that 10-year-olds had significantly higher speaking rates than did 4-year-olds, 13-year-olds had significantly higher speaking rates than did 4- and 7-year-olds, and 16-year-olds and adults both had significantly higher speaking rates than did 10-, 7-, and 4-year-olds. Taken together, speaking rate increases with increased age and with decreased language formulation demands.

The articulation rate of the narratives was compared with the speaking rates of the other tasks. Because the other tasks were found to have no pausing, the speaking rates for the DDK, syllable, and sentence repetitions are also the articulation rates. Similar to speaking rate, a between-groups factorial analysis of variance with follow-up analyses using the LSD minimum mean difference procedure was conducted to examine the effects of age and task on articulation rate. A main effect of age, $F(5, 188) = 14.88$, $p < .001$, was found with articulation rate increasing with age. The main effect of task was significant, $F(3, 188) = 48.90$, $p < .001$. Similar to speaking rate, these main effects were qualified by a significant interaction of task and age, $F(15, 188) = 1.90$, $p < .05$.

Follow-up analyses using the LSD minimum mean difference were used to further examine the Age × Task interaction. For 4-, 10-, 13-, 16-year-olds, and adults, the DDK task had significantly higher articulation rates than the other tasks. Additionally, 13-year-olds demonstrated significantly higher articulation rates for sentences than narratives, and adults demonstrated significantly higher articulation rates for syllables than sentences. For the 7-year-olds, the DDK and syllable repetition task had significantly higher articulation rates than both sentences and narratives.

Focusing on age differences within a task, for the DDK task, adults and 13-year-olds had higher articulation rates than all the other groups. In addition, 4-year-olds had significantly lower articulation rates than 16-, 10-, and 7-year-olds. Syllable repetitions for 7 year-olds were significantly higher in articulation rates than 4-, 10-, 16-year-olds, and adults, and significantly higher for 13-year-olds than 4- and 10-year-olds. For the sentence repetition task, 13-year-olds and adults had significantly higher articulation rates than 4- and 7-year-olds. Similarly 10- and 16-year-olds had significantly higher articulation rates than 4-year-olds. Finally, for the narrative retell task, 16-year-olds and adults had significantly higher articulation rates for 4-, 7-, and 10-year-olds; 7- and 13-year-olds also had significantly higher articulation rates than 4-year-olds. Similar to the speaking rate results, articulation rate also increases with increased age and with decreased language formulation demands.

The change in percent pausing time for narrative retells across age groups is shown in Figure 3. A one-way analysis of variance was conducted to examine the effect of age on percent pausing time. A main effect of age, $F(5, 48) = 9.84$, $p < .001$, was found. Post hoc tests using the LSD procedure using an alpha level of .05 were conducted. The 4- and 7-year-olds had significantly greater percent pausing time than all the other age groups except 10-year-olds. No significant differences were found between 10-, 13-, 16-year-olds, and adults.

Age and Task Effects on Peak Speed

The peak lower lip movement speeds for each speaking task are shown in Figure 4. A between-groups factorial analysis of variance with follow-up analyses was performed to examine the effects of age and task on peak speed. There was no significant interaction between age and task. A main effect of age, $F(5, 188) = 17.66$, $p < .001$, was found. There was also a main effect of task, $F(3, 188) = 17.06$, $p < .001$, with faster speeds for tasks requiring more spoken language production processing demands.

Post hoc tests using the LSD procedure with an alpha level of .05 were performed. Across tasks, 4-year-olds were significantly faster than 13-year-olds. The 16-year-olds were significantly faster than were 4-, 7-, 10-, and 13-year-olds. Adults had significantly faster peak speeds than did 7-, 10-, and 13-year-olds. Across all ages, DDK and syllable repetition tasks were produced with significantly slower peak speeds than the sentence repetition and narrative retell tasks. Overall, 16-year-olds and adults had the fastest peak speeds. Across all participants, tasks requiring greater spoken language production processing demands were produced with faster peak speeds than those requiring less spoken language production processing demands.

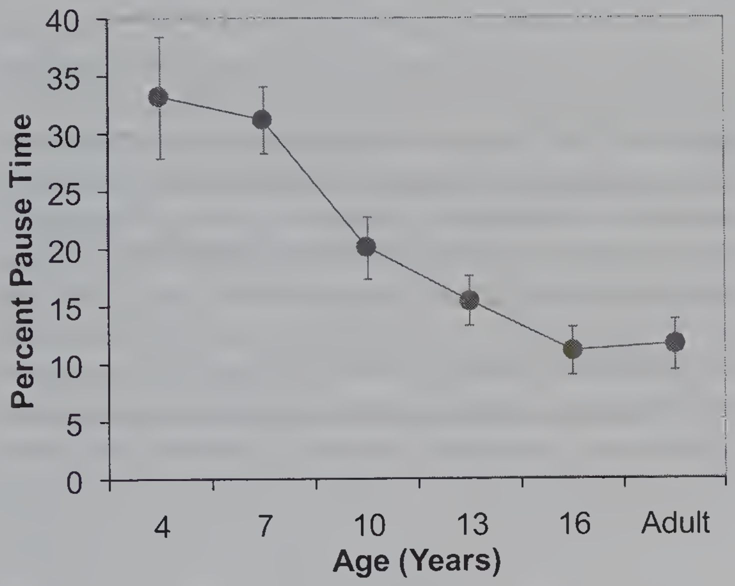


Figure 3. Percent pause time for narrative retell task.

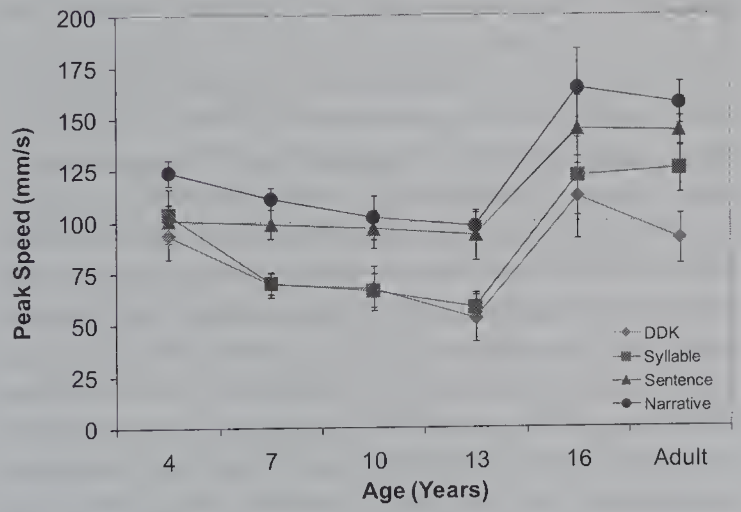


Figure 4. Peak speed of the lower lip marker in mm/s for the diadochokinetic (DDK), syllable repetition, sentence repetition, and narrative retell tasks.

Age and Task Effects on Duration

Duration of the bilabial opening for each speaking task is shown in Figure 5. A between-groups analysis of variance was also conducted on duration. A significant interaction between age and task was also found, $F(15, 188) = 2.12$, $p < .01$, primarily because the narratives had the longest durations except for the 16-year-olds and adult groups for whom syllable repetition had longer durations. Significant main effects for both age, $F(5, 188) = 18.76$, $p < .001$, and task, $F(3, 188) = 105.06$, $p < .001$, were found as well.

Post hoc tests using the LSD minimum mean difference were used to determine how cell means differed from each other. The 4-year-olds had significantly shorter durations for DDK than the

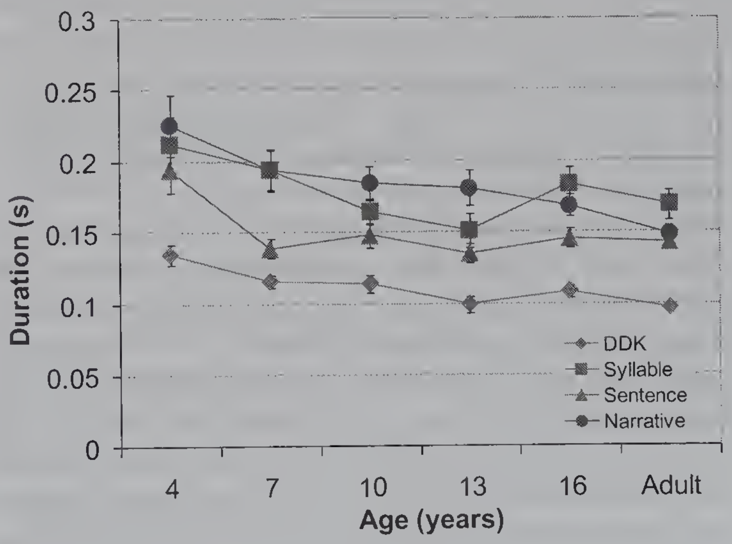


Figure 5. Duration of bilabial opening in seconds for the diadochokinetic (DDK), syllable repetition, sentence repetition, and narrative retell tasks.

other tasks, and sentences had significantly shorter durations than narratives. The 7-year-olds had significantly longer durations for both sentences and DDK when compared to syllables or narratives. The 10-year-olds had significantly shorter durations for DDK than for the other tasks, and narratives showed significantly longer durations than the other tasks. Thirteen-year-olds also had shorter durations for DDK than the other tasks. In addition, narratives had significantly longer durations than syllables or sentences for this age group. Finally, both the 16-year-olds and adults had significantly shorter durations for DDK than other tasks and sentences had significantly shorter durations than syllables.

The LSD minimum mean difference was also used to determine how age groups differed in their duration for each task. In the DDK tasks, 4-year-olds had significantly longer durations than did 10-, 13-, 16-year-olds, and adults. In the syllable repetition task, 4-year-olds had significantly longer durations than all groups except for 7-year-olds, who had significantly longer durations than did 10- or 13-year-olds. In addition, 16-year-olds had longer durations than did 13-year-olds in this task. For the sentence repetition task and the narrative task, 4-year-olds had significantly longer durations than did the other age groups. For narratives, adults had significantly shorter durations than did 7-, 10-, and 13-year-olds. Overall, durations also increased with increased processing for spoken language production and durations generally decreased with age.

Age and Task Effects on Displacement

The displacement of bilabial opening for each speaking task is shown in Figure 6. A between-groups analysis of variance was also conducted on displacement. No significant interaction was found. The main effects of age, $F(5, 188) = 15.14$, $p < .001$, and task, $F(3, 188) = 46.56$, $p < .001$, were significant.

Post hoc tests using the LSD procedure with an alpha level of .05 were performed to examine the main effect of age. Four-year-olds had significantly larger displacements than did the 7-, 10-, 13-year-olds. The 7-year-olds had smaller displacements than did 16-year-olds but larger displacement than did 13-year-olds. Both 10- and 13-year-olds had smaller displacements than did 16-year-olds and adults. Overall, the youngest participants (4-year-olds) and the oldest participants (16-year-olds and adults) produced larger displacements for bilabial opening than did 7-, 10-, and 13-year-olds.

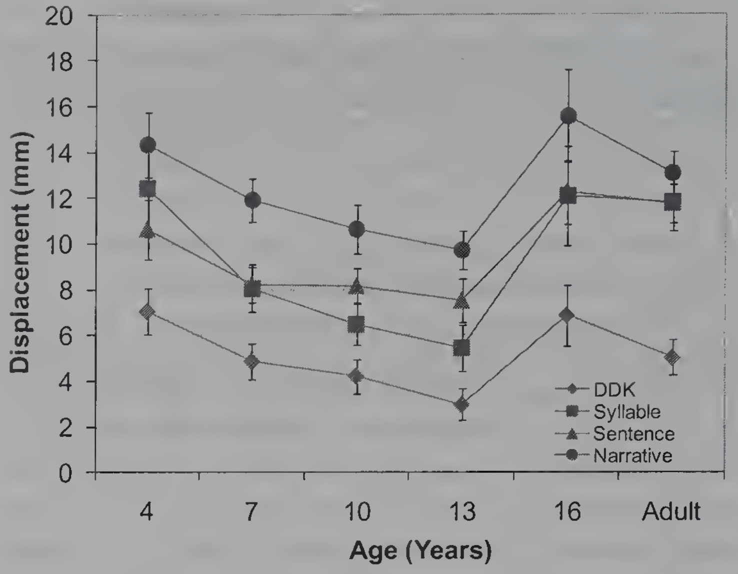


Figure 6. Displacement of the lower lip marker for bilabial opening in millimeters for the diadochokinetic (DDK), syllable repetition, sentence repetition, and narrative retell tasks.

Post hoc tests using the LSD procedure using an alpha level of .05 were performed to examine the main effect of task. For all age groups the DDK task had significantly smaller displacements as compared to other tasks. Both the syllable repetition and sentence repetition tasks had significantly smaller displacements as compared to the narrative retell task. There was no significant difference in displacement between the syllable and sentence repetition tasks. Generally, displacements increased with tasks requiring greater spoken language processing.

Discussion

The current study investigated developmental changes in speaking rate and articulatory kinematics in speaking tasks varying in task complexity. The data suggest that speaking rate increases gradually with age; however, movement speed and displacement decrease from preschool to 13 years of age before sharply increasing. These findings suggest that the development of speaking rate can be characterized by two major phases. The first phase occurred between 4 and 13 years when the greatest increases in speaking rate occurred. This change was clearly driven by decreases in movement displacement for the DDK, syllable, and sentence tasks and by decreases in movement displacement and pause time for the narrative task, and not increases in movement speed. The second phase occurred between 13 and 16 years of age when increases in speaking rate were primarily driven by increases in both peak speed and displacement with concomitant decreases in pauses time. The observation of gradual increases in speaking rate across the ages was striking given that it occurred in the context of more abrupt, underlying changes in speech motor control, specifically, the nonmonotonic increases in peak speed and displacement. One interpretation of these findings is that they demonstrate how, over time, children adapt to significant changes in their vocal tract anatomy to make gradual progress toward achieving adult-like speech performance. Taken together, the findings also suggest that increases in speaking rate are achieved through efficiencies in motor control and processing that support spoken language production. In contrast, evidence for the contribution of biologically-driven aspects (i.e., movement speed) was minimal with one exception being the major shift in articulatory kinematics that occurred between 13 and 16 years of age, which is during a period of accelerated vocal tract growth.

All of the variables measured in this study were affected by task demands; however, the demands of spoken language production produced opposite effects on speaking rate and articulatory kinematic variables. Specifically, across all age groups, speaking rate was faster for tasks with less spoken language production demands, such as DDK, and slower for tasks with greater spoken language production demands, such as sentences and narrative retell. Peak speed, duration, and displacement showed the opposite trend, with less demanding tasks produced with slower peak speeds, shorter durations, and smaller displacements than were more demanding tasks. One interpretation of these findings is that in young talkers' articulator movements become less efficient as the demands for spoken language increase.

Developmental Trends

Speaking and articulation rate changes with age. Speaking and articulation rate for all speaking tasks increased gradually from 4 to 13 years of age, replicating previous findings in the literature (e.g., Amster, 1984; Haselager et al., 1991; Kowal et al., 1975; Walker et al., 1992). Speaking rate did not change after 13 years of age, replicating earlier findings that 12-year-old children speak at approximately 90% of the adult speaking rate (Walsh & Smith, 2002). However, a nonsignificant decrease in speaking rate was observed between 13 and 16 years of age for the DDK, syllable repetition, and sentence repetition tasks. This nonmonotonic trend in speaking rate is similar to previously reported findings that developmental changes in speaking rate are not necessarily linear. For example, Kowal et al. (1975) found that although 14- and 18-year-olds had very similar speaking rates, a transient increase in speaking rate was observed between the ages of 14 and 16 years. In addition, earlier studies of speaking rate have observed a plateau in speaking rate between the ages of 3 and 5 years (Pindzola, Jenkins, & Lokken, 1989) and a transient decrease in speaking rate at the age of 5 years (Walker & Archibald, 2006). In the current investigation, the percentage of pausing time during narrative retell decreased with age. This finding is similar to previous findings that demonstrate the percentage of pausing decreased from the ages of 4 to 16 years (Haselager et al., 1991; Kowal et al., 1975).

The duration of the speech movements, or the time it took for participants to produce a bilabial opening, indicated the net effect of the age-related changes observed in articulatory speed and displacement. Duration steadily decreased with age, but was significantly longer for 4- and 7-year-olds than it was for the older groups. Duration decreased as speaking rates increased. This finding is similar to those reported in earlier studies demonstrating that sentence durations decrease from 7 years of age to adults, suggesting that speech segments are produced with shorter durations with age (Chermak & Schneiderman, 1985). This decrease in pausing and duration may represent a decrease in the time needed to: (a) conceptualize and formulate an utterance (Kowal et al., 1975; Rochester, 1973) and (b) motorically plan an utterance as increased motor practice has been shown to decrease planning time (Oytam, Neilson, & O'Dwyer, 2005).

Biologic Factors Did Not Have an Obvious Influence on the Development of Speaking Rate

Unlike previous studies, this investigation examined how speaking rate development changes in the context of underlying articulatory kinematics. Although speaking rate increased incrementally between 4 and 13 years of age, the kinematic measures show a very different developmental course. Peak speed of articulatory movement decreased from 4 to 13 years before dramatically increasing between the ages of 13 to 16 years, demonstrating a strong disassociation between speaking rate and peak speed development. This finding suggests that peak speed may play only a minor role in limiting the rate at which children speak.

Age-related changes in movement displacement closely parallel changes in movement speed. Generally, 4-year-olds were not significantly different in displacements as compared to the older participants (16-year-olds and adults) despite their smaller anatomical structures; this finding replicates earlier studies that demonstrate 5-year-olds do not have smaller displacements than adults during speech (Riely & Smith, 2003). However, Riely and Smith (2003) only examined two age groups: 5-year-olds and adults; the current study examined age groups between 4 years and adulthood. In the current investigation, displacement significantly decreased from the ages of 4–7 years of age, then plateaued between 7 and 13 years of age, before increasing again at 16 years of age. The similar displacements among the 7-, 10-, and 13-year-olds suggest that the ability to scale oral movements has a protracted developmental course. Potentially, this may reflect refinement in the scaling of articulatory displacements, similar to findings in the limb literature for grasping (e.g., Potter et al., 2006). Another possible reason for some of the age-related changes in the displacements may be anatomical growth. Previous research has demonstrated a major growth spurt in the vocal tract between the ages of 13 and 16 years (Fitch & Giedd, 1999; Vorperian et al., 2009) and may explain the large increase in articulatory displacements found between these ages.

Task Effects

Speaking rate decreased with increasing production demands. Across all age groups, speaking rate slowed predictably in response to increasing task demands. In the current study, the DDKs, which had the least demands on cognitive, linguistic, and motor processing, were produced with faster speaking rates than were the sentence repetitions. Sentences, which were presumed to impose the greater spoken language production processing demands than DDK but less than the narratives, were produced with faster speaking rates than were the narrative retell tasks. Similar task effects on speaking rate have been reported in prior research on speech development. For instance, DDK speaking rates have been found to be significantly faster than spontaneous speech across childhood (Haselager et al., 1991). In addition, speaking material that is highly familiar, such nursery rhymes, and therefore presumed to require less cognition or language formulation resources, are produced at significantly faster speaking rates than less familiar sentences in preschoolers (Walker & Archibald, 2006). Adults show similar trends in speaking rate changes with speaking task; speaking tasks that are perceived as more difficult are associated with slower speaking rates, primarily due to more frequent and longer pauses (Mitchell et al., 1996).

Speaking rates for the syllable repetition task were less predictable than were speaking rates observed for the other tasks. Specifically, the syllable repetition task was presumed to have required more processing than the DDK task but less than the sentence repetition task. Therefore, the rate of speech during this task was predicted to fall between the DDK and sentence repetition tasks. In the current study, the speaking rate for the syllable repetition decreased with age, in relation to the other tasks. For instance, syllables are produced quite quickly in relation to the other tasks for younger participants but older participants, such as the 16-year-olds and the adults produced them at the same rate as the narrative retell. One interpretation of this finding is that the younger participants processed the syllables like a nonlinguistic task such as DDK, whereas adults processed the syllables more like meaningful speech. The reason for this putative change in processing is not obvious but could be related to changes in how speech representations are affected by emerging semantic, lexical, phonologic, and even literacy skills (Munson, Swenson, & Manthei, 2005; Stoel-Gammon, 2011).

Peak speed, duration, and displacement. A consistent task effect was also seen in the peak articulatory speed data. In contrast to the speaking rate data, less demanding speaking tasks (i.e., DDK and syllable repetition) were produced with significantly slower movement speeds than were more demanding speaking tasks (i.e., sentences and narratives). This somewhat paradoxical result (speed might be expected to be faster during low-demand task than during high-demand task) suggests that factors other than processing demands govern the speed at which talkers move their articulators. One possible explanation for this finding is that lip displacements were larger for the high-demand tasks than for the low-demand tasks. It has long been known that large displacements are produced at faster speed than small displacements (see Ostry, Keller, & Parush, 1983). The observation that lip displacements become larger during the more complex tasks suggests that articulator movements become less efficient as the demands for spoken language increase. An alternative explanation is that the more demanding tasks were more distinctly or hyperarticulated (Lindblom, 1990) because they contained linguistically relevant information. In contrast, talkers may have hypoarticulated the simple stimuli (i.e., truncate displacement) because they were willing to underspecify speech sounds because the goal of the task was to produce the syllables as rapidly as possible without having to convey linguistic information. Prior research has consistently demonstrated that articulatory displacements are truncated during rapidly produced syllables or speech (Mefferd & Green, 2010; Westbury & Dembowski, 1993). In short, if the achievement of adult-like speaking rates primary involves decreasing articulatory displacements, it is not surprising to see that articulatory movement speeds decrease proportionately.

The task-related findings on movement speed are also consistent with Lindblom's (1990) concept of listener-oriented articulatory control, where the specification of articulatory movements varies depending on the speaker's perception of the listeners' demands. Thus, for the sentence repetition and the narrative retell tasks, talkers may have increased the extent of their articulatory movements to enhance speech clarity for the purpose of conveying linguistic information.

Similarly, task effects were observed in the duration of lip and jaw opening, with duration generally decreasing with decreasing processing demands. Surprisingly, the older participants (16-year-olds and adults) produced syllables with longer durations than the narratives. Previous research has demonstrated that a shift in motor planning may occur as children increase cognitive and linguistic processing. Older children and adults decrease the duration of some phonemes and syllables when a simple (SVO) sentence is embedded within a complex sentence, similar to adult speakers; however, younger children do not show any differences in segmental duration between simple and complex sentences (Sadagopan & Smith, 2008). Sadagopan and Smith (2008) suggest that this change in duration may be the result of increased linguistic maturity and flexibility that allows older talkers to plan longer sequences at a time rather than using word-by-word or syllable-by-syllable strategy. This hypothesis may account for the difference in task effects for the 16-year-olds and the adults as compared to the other age groups. The older participants have the flexibility to decrease the segmental duration for the /b/ in order to produce longer utterances in the narrative retell; however, longer durations are utilized for shorter utterances such as simple syllables, perhaps to allow for greater accuracy in producing the phonetic targets. Similarly, the hypothesis put forth by Sadagopan and Smith (2008) may also account for the high degree of similarity of the durations for bilabial openings for the syllable, sentence, and narrative retell task of the younger group because this group lacks the ability to motorically plan units larger than a syllable or a word.

Conclusion

The current investigation was designed to test several hypotheses regarding the roles of biologic factors, and motor learning and spoken language processing gains on the typical development of speaking rate. Speaking rate increased between 4 and 13 years of age and consistent task effects were seen across all ages. More complex tasks were produced at slower speaking rates than the less complex tasks. Pausing data from the narratives suggest that children become more efficient in cognitive, linguistic, and motor planning. Consistent task effects were also seen for kinematic measures; however, the direction of the task effects for speaking rate and articulatory speed were in the opposite directions; less complex tasks had faster speaking rates but slower peak speeds as compared to more complex tasks. This effect may represent talkers' attempts to meet the demands of each task in the most efficient manner.

References

American Speech-Language-Hearing Association Panel on Audiologic Assessment. (1997). *Guidelines for audiologic screening*. Rockville, MD: American Speech-Language-Hearing Association.

Amster, B. J. (1984). *The rate of speech of normal preschool children (Articulatory development, fluency, language, stuttering)*. Unpublished doctoral dissertation, Temple University, Philadelphia, PA.

Barlow, S. M., Finan, D. S., Bradford, P. T., & Andreatta, R. D. (1993). Transitional properties of the mechanically evoked perioral reflex from infancy through adulthood. *Brain and Language, 623*, 181–188.

Butterworth, S. (1930). On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer, 7*, 536–541.

Campbell, T. F., & Dollaghan, C. A. (1995). Speaking rate, articulatory speed, and linguistic processing in children and adolescents with severe traumatic brain injury. *Journal of Speech and Hearing Research, 38*, 864–875.

Chermak, G. D., & Schneiderman, C. R. (1985). Speech timing variability of children and adults. *Journal of Phonetics, 13*, 477–480.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dromey, C., & Benson, A. (2003). Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance. *Journal of Speech, Language and Hearing Research, 46*, 1234–1246.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of Acoustic Society of America, 106*, 1511–1522.

Flipsen, P., Jr. (2002). Longitudinal changes in articulation rate and phonetic phrase length in children with speech delay. *Journal of Speech, Language and Hearing Research, 45*, 100–110.

Flipsen, P., Jr. (2003). Articulation rate and speech-sound normalization failure. *Journal of Speech, Language and Hearing Research, 46*, 724–737.

Goffman, L., & Smith, A. (1999). Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 649–660.

Goldfield, E. C., Kay, B. A., & Warren, W. H. (1993). Infant bouncing: The assembly and tuning of action systems. *Child Development, 64*, 1128–1142.

Green, J. R., Beukelman, D. R., & Ball, L. J. (2004). Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. *Journal of Medical Speech-Language Pathology, 12*, 149–154.

Green, J. R., & Nip, I. S. B. (2010). Some organization principles in early speech development. In B. Maassen & P. H. H. M. Van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 171--188). Oxford, UK: Oxford University Press.

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant directed speech. *Journal of Speech, Language, and Hearing Research, 53*, 1529–1542.

Green, J. R., & Wilson, E. M. (2006). Spontaneous facial motility in infants: A 3D kinematic analysis. *Developmental Psychobiology, 48*, 16--28.

Greene, J. O. (1984). Speech preparation processes and verbal fluency. *Human Communication Research, 11*, 61–84. doi:10.1111/j.1468-2958.1984.tb00038.x

Greene, J. O., & Cappella, J. N. (1986). Cognition and talk: The relationship of semantic units to temporal patterns of fluency in spontaneous speech. *Language and Speech, 29*, 141–157.

Hall, K. D., Amir, O., & Yairi, E. (1999). A longitudinal investigation of speaking rate in preschool children who stutter. *Journal of Speech, Language and Hearing Research, 42*, 1367–1377.

Haselager, G. J. T., Slis, I. H., & Rietveld, A. C. M. (1991). An alternative method of studying the development of speech rate. *Clinical Linguistics & Phonetics, 5*, 53–63.

Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. Oxford: Oxford University Press.

Kelly, E. M., & Conture, E. G. (1992). Speaking rates, response time, latencies, and interrupting behaviors of young stutterers, nonstutterers, and their mothers. *Journal of Speech and Hearing Research, 35*, 1256–1267.

Kowal, S., O'Connell, D. C., & Sabin, E. J. (1975). Development of temporal patterning and vocal hesitations in spontaneous speech. *Journal of Psycholinguistic Research, 4*, 195–207.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Netherlands: Kluwer Academic.

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53*, 1206–1219.

Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. *Journal of Speech and Hearing Research, 39*, 93–104.

Müller, K., & Hömberg, V. (1992). Development of speed of repetitive movements in children is determined by structural changes in corticospinal efferents. *Neuroscience Letters, 144*, 57–60.

Müller, K., Hömberg, V., & Lenard, H. G. (1991). Magnetic stimulation of motor cortex and nerve roots in children: Maturation of cortico-motoneuronal projections. *Electroencephalography and Clinical Neurophysiology, 8*, 163–170.

Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech, Language and Hearing Research, 48*, 108–124.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology. Human Perception and Performance, 9*, 622–636.

Oytam, Y., Neilson, P. D., & O'Dwyer, N. J. (2005). Degrees of freedom and motor planning in purposive movement. *Human Movement Science, 24*, 710–730.

Pindzola, R. H., Jenkins, M. M., & Lokken, K. J. (1989). Speaking rates of young children. *Language, Speech, and Hearing Services in Schools, 20*, 133–138.

Potter, N. L., Kent, R. D., Lindstrom, M. J., & Lazarus, J. C. (2006). Power and precision grip force control in three-to-five-year-old children: Velocity control precedes amplitude control in development. *Experimental Brain Research, 172*, 246–260.

Riely, R. R., & Smith, A. (2003). Speech movements do not scale by orofacial structure size. *Journal of Applied Physiology, 94*, 2119–2126.

Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research, 2*, 51–81.

Ryan, B. (2000). Speaking rate, conversational speech acts, interruption, and linguistic complexity of 20 pre-school stuttering and non-stuttering children and their mothers. *Clinical Linguistics & Phonetics, 14*, 25–51.

Sadagopan, N., & Smith, A. (2008). Developmental changes in the effects of utterance length and complexity on speech movement variability. *Journal of Speech, Language and Hearing Research*, *51*, 1138–1151. doi:10.1044/1092-4388(2008/06-0222)

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals screening test* (4th ed.). San Antonio, TX: Psychology Corporation.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language and Hearing Research*, *41*, 18–30.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, *45*, 22–33.

Smith, B. L., & Gartenberg, T. E. (1984). Initial observations concerning developmental characteristics of labiomandibular kinematics. *Journal of Acoustic Society of America*, *75*, 1599–1605.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*, 1–34. doi:10.1017/S0305000910000425

Sturm, J. A., & Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Language, Speech, and Hearing Services in Schools*, *38*, 47–59. doi:10.1044/0161-1461(2007/005)

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., et al. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of the Acoustical Society of America*, *125*, 1666–1678. doi:10.1121/1.3075589

Walker, J. F., & Archibald, L. M. D. (2006). Articulation rate in preschool children: A 3-year longitudinal study. *International Journal of Language & Communication Disorders*, *41*, 541–565.

Walker, J. F., Archibald, L. M. D., Cherniak, S. R., & Fish, V. G. (1992). Articulation rate in 3- and 5-year-old children. *Journal of Speech, Language and Hearing Research*, *35*, 4–13.

Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research*, *45*, 1119–1133.

Westbury, J. R., & Dembowski, J. (1993). Articulatory kinematics of normal diadochokinetic performance. *Annual Bulletin–Research Institude of Logopedics and Phoniatrics*, *27*, 13–36.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals–Preschool* (2nd ed.). San Antonio, TX: Psychology Corporation.

Ziegler, W. (2002). Task-related factors in oral motor control: Speech and oral diadochokinesis in dysarthria and apraxia of speech. *Brain and Language*, *80*, 556–575. doi:10.1006/brln.2001.2614

Child Development, July/August 2013, Volume 84, Number 4, Pages 1338–1354

Associations Among Pubertal Development, Empathic Ability, and Neural Responses While Witnessing Peer Rejection in Adolescence

Carrie L. Masten
Vanderbilt University

Naomi I. Eisenberger
University of California, Los Angeles

Jennifer H. Pfeifer
University of Oregon

Natalie L. Colich
Stanford University

Mirella Dapretto
University of California, Los Angeles

Links among concurrent and longitudinal changes in pubertal development and empathic ability from ages 10 to 13 and neural responses while witnessing peer rejection at age 13 were examined in 16 participants. More advanced pubertal development at age 13, and greater longitudinal increases in pubertal development, related to increased activity in regions underlying cognitive aspects of empathy. Likewise, at age 13 greater perspective taking related to activity in cognitive empathy-related regions; however, affective components of empathy (empathic concern and personal distress) were associated with activity in both cognitive and affective pain-related regions. Longitudinal increases in empathic ability related to cognitive and affective empathy-related circuitry. Findings provide preliminary evidence that physical and cognitive-emotional development relate to adolescents' neural responses when witnessing peer rejection.

In adolescence, peer relationships take on new importance as youth spend more time with peers and place more value on peer belonging (Brown, 2004). Unfortunately, due to the high value placed on maintaining peer relationships, peer rejection and bullying become increasingly prevalent at this age (Brown, 2004), and are associated with many negative outcomes (Arseneault, Bowes, & Shakoor, 2010; Brendgen & Vitaro, 2008; Isaacs, Hodges, & Salmivalli, 2008; Lev-Wiesel, Nuttman-Shwartz, & Sternberg, 2006; Prinstein, Sheah, & Guyer, 2005). Furthermore, it is not just increases in firsthand experiences of peer rejection that adolescents worry about. Even adolescents who are not themselves rejected by others are still exposed to, and impacted by, the peer rejection they see happening to others (Janes & Olson, 2000; Nishina & Juvonen, 2005). In fact, most adolescents witness peer rejection on a regular basis in their daily environment (Rivers, Poteat, Noret, & Ashurst, 2009) and observe peer rejection and bullying at least as often as they experience it firsthand (Nishina & Juvonen, 2005).

Recently, neuroimaging techniques have been used to examine adolescents' emotional responses and empathy when they witness others being rejected (Masten, Eisenberger, Pfeifer, & Dapretto, 2010, 2012). Functional magnetic resonance imaging (fMRI) is particularly useful because it permits the

---

This work was supported by the Santa Fe Institute Consortium, as well as by an Elizabeth Munsterberg Koppitz Award and a Ruth L. Kirschstein National Research Service Award to C. Masten. For generous support, the authors also wish to thank the Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, Ahmanson Foundation, Tamkin Foundation, Jennifer Jones-Simon Foundation, Capital Group Companies Charitable Foundation, Robson Family, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, and Northstar Fund. This project was in part also supported by grants (RR12169, RR13642, and RR00865) from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH); its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCR or NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Correspondence concerning this article should be addressed to Carrie L. Masten, Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Peabody #552, Nashville, TN 37203. Electronic mail may be sent to carrie.masten@gmail.com.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0017
DOI: 10.1111/cdev.12056

study of processes underlying peer interactions as they are occurring, rather than relying solely on retrospective self-reports or observer ratings, which are limited in their ability to tap in-the-moment processes. In addition, fMRI has the potential to disentangle simultaneous, distinct processes that co-occur when a social interaction is being observed, which together create an individuals' overall subjective experience. Furthermore, fMRI is also particularly useful for studying social developmental processes because it allows examination of how social contexts or individual differences might alter underlying affective and cognitive processes. In other words, it can help reveal how certain developmental constructs (e.g., pubertal development or empathic processing) might modulate affective and cognitive neural processes engaged during peer-related events. Thus, this method provides a unique perspective when trying to understand correlates of adolescent behavior (see also Masten & Eisenberger, 2009; Pfeifer & Blakemore, 2012).

Prior work has shown that affective and cognitive neural processes may impact how individuals subjectively experience the observation of peer exclusion (details in the next section). However, as major social changes in adolescence (i.e., increased peer rejection) coincide with many physical and cognitive changes (see Nelson, Leibenluft, McClure, & Pine, 2005; Steinberg, 2004, 2007), neural processes underlying peer interactions are likely related to ongoing physical and cognitive emotional development across this transitional period. For example, structural and functional reorganization of neural networks involved in social and empathic processes occur during puberty and are thought to interact with social and behavioral changes at this age (Blakemore, 2008; Blakemore, Burnett, & Dahl, 2010; Bramen et al., 2011; Decety, 2010; Nelson et al., 2005). Moreover, hormonal changes precipitated by puberty are thought to be one of the most important influences driving social reorientation in adolescence (e.g., Forbes & Dahl, 2010) and are therefore likely to have direct links with the neural underpinnings of adolescent social processes. Thus, the goal of this study was to build on prior research by examining how longitudinal changes across the transition to adolescence in two developmental indices—pubertal development, an indicator of physical change, and empathic ability, an indicator of both cognitive (i.e., perspective taking) and emotional (i.e., empathic concern, personal distress) change—relate to adolescents' neural responses when they witness peer rejection.

The Cognitive and Affective Components of Empathy

Research examining the processes underlying empathy has typically focused on two different aspects of empathic experiences (e.g., Baron-Cohen, 2003; Eisenberg, 2000; Eisenberg, Spinrad, & Sadovsky, 2006). First, there are the cognitive abilities that enable people to understand the mental states of other (skills often termed *mentalizing*; Frith, Leslie, & Morton, 1991). Second, there is the affective dimension of empathy, which allows people to share or "mirror" the emotions and pain of others. Recently, neuroimaging research has indicated that two different neural networks underlie the cognitive and affective components of empathy (Decety, 2010; Decety & Meyer, 2008; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009; Singer, 2006; Singer & Lamm, 2009). Typically, the cognitive component of empathy is thought to rely on the "mentalizing network"—a network of regions that are associated with various aspects of mentalizing, including the posterior superior temporal sulcus (pSTS), the temp oral poles, the precuneus and posterior cingulate cortex (PCC), the temporal–parietal junction (TPJ), and the medial and dorsomedial prefrontal cortex (MPFC–DMPFC; Frith & Frith, 1999, 2003, 2006; Mitchell, Banaji, & Macrae, 2005; Saxe, 2006; Singer, 2006). In contrast, the neural regions underlying the affective component of empathy are thought to be those that are activated during both firsthand and vicarious affective experiences. For example, the anterior insula is activated by both direct and observed experiences of disgust (Wicker et al., 2003), the amygdala is activated during direct and observed fear (Whalen et al., 2001), and the dACC and anterior insula are activated by direct and observed physical pain in adults (Botvinick et al., 2005; Jackson, Bruney, Meltzoff, & Decety, 2005; Morrison, Lloyd, Di Pellegrino, & Roberts, 2004; Singer et al., 2004; Singer et al., 2006) and children (Decety, Michalska, & Akitsuki, 2008). Both the insula and amygdala are also activated among children when they both observe and make emotional face expressions (Pfeifer, Iacoboni, Mazziotta, & Dapretto, 2008).

Neural Correlates of Witnessing Social Exclusion

A few studies have examined the neural correlates of empathy-related processes engaged while observing social exclusion specifically. In adults, viewing exclusion (vs. inclusion) activates regions in the mentalizing network (DMPFC, MPFC, VMPFC, precuneus; Masten, Morelli, & Eisenberger, 2011). The affective pain network (dACC, AI) is also activated during observed exclusion in individuals reporting high levels of empathy (Masten et al., 2011), and among friends of the excluded victim (Beeney, Franklin, Levy, & Adams, 2011; Meyer et al., 2012)—suggesting that sharing the affective experience of the victim might depend on feelings of closeness or similarity to the victim or being able to relate to him or her more. In terms of adolescents, a recent study demonstrated that 13-year-olds also display activity in the mentalizing network (DMPFC, MPFC, precuneus, pSTS) when viewing peer exclusion (Masten et al., 2010). Interestingly, in this study, there was no evidence of affective pain processes even in adolescents who reported being more empathic. However, we used a single unidimensional empathy measure in this early study, which may not have been sensitive enough to separately index cognitive and affective components of empathy, or their respective associations with distinct networks of brain activity.

Pubertal Development and Changes in Empathic Ability in Adolescence

To understand adolescents' neural responses when they witness negative interactions among their peers, it is crucial to consider the many developmental factors that might modulate individuals' responses. First, it is commonly hypothesized that many of the social changes occurring in adolescence—including the increasing salience of peers—coincide with early stages of pubertal development, and occur as a result of puberty-related hormonal changes (Blakemore et al., 2010; Forbes & Dahl, 2010; Nelson et al., 2005; Steinberg, 2004, 2007). Next, youth also make significant strides in terms of empathic ability as they transition to adolescence (Eisenberg, Miller, Shell, McNally, & Shea, 1991), which is not surprising given increased time spent with peers and concern about peer acceptance. In fact, given known links between empathy and successful social communication (Denham, Renwick-DeBardi, & Hewes, 1994; de Vignemont & Singer, 2006; Zahn-Waxler, Cole, Welsh, & Fox, 1995; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992), empathic ability may continue to develop in adolescence precisely because of the increased need for empathy-related skills that facilitate successful peer interactions (Eisenberg & Morris, 2004; Eisenberg, Morris, McDonald, & Spinrad, 2009). Although there has been a shortage of empirical work examining direct links between puberty and empathy, recent research has linked pubertal development with: assessment of others' opinions and feelings (Burnett, Thompson, Bird, & Blakemore, 2010)—a skill that becomes particularly important in adolescence when the perceived opinions of peers are highly salient (Brown, 2004); heightened stress due to social evaluation (Gunnar, Wewerka, Frenn, Long, & Griggs, 2009; Sumter, Bokhorst, Miers, Van Pelt, & Westenberg, 2010); increased sensation seeking (Martin et al., 2002) and emotional reactivity (Silk et al., 2009; Spear, 2009); as well as greater "mentalizing" ability (Keulers, Evers, Stiers, & Jolles, 2010). It is likely also associated with increasing interest in, and efforts to interact with, peers (Forbes & Dahl, 2010). In fact, Forbes and Dahl (2010) posit that puberty-related hormonal increases are a major catalyst underlying adolescents' ability and motivation to maintain peer relations. Thus, it is crucial to examine how pubertal factors relate to peer interactions and social cognitive advances at this age.

In terms of brain development, a significant degree of structural reorganization and functional changes are thought to occur concurrently with puberty and relate to the social changes that accompany adolescence (Blakemore, 2008; Blakemore et al., 2010; Bramen et al., 2011; Nelson et al., 2005). Similarly, many of the regions linked to both the cognitive and affective components of empathy undergo significant structural and functional change across the adolescent transition (Blakemore, 2008; Blakemore et al., 2010; Decety, 2010; Nelson et al., 2005). To our knowledge, interrelated influences of puberty and empathy on neural processing have not yet been examined. However, pubertal development has been linked with prefrontal cortical activity during reward processing (Forbes et al., 2010), as well as both limbic and prefrontal cortical activity during facial emotion processing (Moore et al., 2012). Thus, pubertal development appears to be associated with activity in some of the same neural networks that are involved in both cognitive and affective aspects of empathy. Together, this research indicates that pubertal development and increasing empathic ability likely have important, interrelated links with adolescents' social experiences at behavioral and neural levels.

Clearly, understanding the roles of pubertal development and empathic ability is crucial to understanding adolescents' social interactions. Moreover, witnessing peer rejection is a particularly common type of peer interaction in adolescence that depends on relatively advanced cognitive abilities and higher order cortical function for its understanding. Thus, it is important to examine how pubertal development and empathic ability relate to adolescents' responses when they witness these peer encounters. So far, there is limited research on this topic—particularly with regard to how these constructs are represented at the neural level. Additionally, little is known about how changes in pubertal development and empathic ability across the transition to adolescence are associated with neural functioning in the context of witnessing peer rejection.

The goal of this study was to examine how two developmental indices—pubertal development and empathic ability, measured both concurrently (at age 13) and longitudinally across the adolescent transition (from ages 10 to 13)—might be related to adolescent's neural responses when they witnessed peer rejection (i.e., a neural index of their empathy for peers) at age 13. In particular, we measured pubertal development via self-reports of physical appearance and maturation. In addition, we used three separate dimensions of empathic ability—empathic concern, personal distress, and perspective taking (instead of the single dimension examined in Masten et al., 2010)—because dimensions of empathy that are more cognitively or affectively focused might differentially relate to activity in mentalizing or affective pain regions. We also measured *concurrent* self-reports, as well as *longitudinal increases* across the adolescent transition for both pubertal development and empathic ability, to examine: (a) current status (i.e., the level of development achieved) and (b) changes (i.e., the amount of development occurring during the transition) for each of these measures, in relation to neural responses to peers.

Measuring longitudinal increases was of particular interest given recent literature focusing on *velocity* of change as a key predictor of social emotional outcomes in adolescence. For example, there is wide variation in the rate at which youth progress through puberty (e.g., Mendle, Harden, Brooks-Gunn, & Graber, 2010), and because puberty is itself a stressful life transition that all youth must face (see Ge et al., 2003), physiological or cognitive changes occurring at a faster than average rate might result in a need for especially rapid adjustment to new social demands and experiences. Indeed, accelerated pubertal change or "pubertal tempo" (i.e., a larger degree of pubertal maturation across a specific time frame) has been linked to depressive symptoms in adolescence, particularly for boys (Ge et al., 2003; Mendle et al., 2010).

Hypotheses

First, based on previous developmental research, we predicted that pubertal development and empathic ability would be positively associated with each other both at age 13 and across the adolescent transition (from ages 10 to 13).

In terms of neural associations, we hypothesized that adolescents who reported more advanced pubertal development at age 13 would display greater activity in regions associated with mentalizing and affective pain processing while seeing another adolescent being excluded. In addition, we explored whether longitudinal increases in pubertal development from ages 10 to 13 might also relate to greater neural responses suggestive of empathy for an excluded peer (in both mentalizing and affective pain-related regions) at age 13. Although prior work on this topic is limited, we speculated that rapid physical change coinciding with the social contextual changes that characterize the adolescent transition (e.g., greater peer orientation, middle school transition) might relate to enhanced sensitivity to peers' emotions. In other words, because emotional advances are known to accompany pubertal change (Burnett et al., 2010; Forbes & Dahl, 2010; Keulers et al., 2010), one possibility is that rapid pubertal development—during the precise period when peer importance increases—might relate to similarly accelerated increases in peer salience and emotional understanding.

In terms of empathic ability, we expected that greater empathic ability at age 13 would, in general, relate to more cognitive and affective empathy-related neural activity during observed peer exclusion. Specifically, however, we expected that higher perspective taking scores would relate to more neural evidence of mentalizing, while empathic concern and personal distress would relate to more neural evidence of affective pain processing. Regarding longitudinal changes in empathic ability, we expected positive associations between increasing empathic ability from ages 10 to 13 and neural evidence of both mentalizing and affective pain processes at age 13. But, as the high salience of peers at this age might increase anxiety about negative peer interactions (i.e., more personal distress), or more concern for others who are observed being negatively treated (i.e., greater empathic concern), we expected that increases in these affective dimensions of empathic ability in particular might relate to corresponding increases in neural activity related to affective components of empathy during observed peer exclusion.

Finally, given that pubertal changes, increasing social cognitive skills, and neural functioning during peer interactions are likely to be meaningfully interrelated during adolescence, exploratory mediation analyses were conducted to derive clues about the possible directionality of the relations among these variables. Specifically, we tested two mediation pathways. First, given the commonly posited theory that functional reorganization of the brain accompanies puberty and relates to subsequent social changes (Blakemore, 2008; Blakemore et al., 2010; Bramen et al., 2011; Nelson et al., 2005), we examined if a neural mechanism (increased responses during observed exclusion) might help explain the positive relation between pubertal development and empathic ability. Next, as emotional and cognitive advances are known to accompany puberty and relate to emerging patterns of brain function (Burnett et al., 2010; Forbes & Dahl, 2010; Keulers et al., 2010), we examined whether a cognitive-emotional mechanism (increasing empathic abilities) might help explain the positive relation between pubertal development and neural responses to observed peer exclusion.

Method

Participants

An ethnically and socioeconomically diverse sample of 16 typically developing youth (nine females) completed data collection at two time points separated by 3 years ($M = 3.17$ years, $SD = .29$). To examine the effect of pubertal status unconfounded by age, we recruited participants in an extremely narrow age range. At Time Point 1, participants were 10 years old ($M = 10.07$ years old, $SD = .30$), and at Time Point 2, they were 13 years old ($M = 13.17$ years old, $SD = .29$) and had attended at least 1 year of middle school. Participants came from a range of ethnic backgrounds—62% Caucasian, 19% Asian, 13% Latino (including one individual who reported being Latino and Native American), 6% African American—and socioeconomic backgrounds; maternal education ranged from high school diploma to advanced graduate degrees (*Mdn* = bachelor's degree), and total household income ranged from less than \$25,000 to greater than \$400,000 (*Mdn* = \$65,000–\$80,000). The study was initially presented to participants and their parents as a longitudinal examination of adolescent brain development, and recruitment was performed via mass mailings, summer camps, and fliers distributed in the Greater Los Angeles area so as to obtain a sample that was as representative as possible. Participants had no history of psychiatric, neurological, or medical disorders. All participants and parents provided assent or consent, which was approved by UCLA's Institutional Review Board.

This sample is identical to that examined (at age 13 only) in Masten et al. (2010), but all analyses and findings are unique from those reported previously.

Overview of Procedures and Task

The goal of this study was to examine how participants' neural responses to witnessing peer rejection related to their concurrent pubertal development and empathic ability, as well as their longitudinal changes in these variables across the 3-year period preceding the witnessed event (i.e., from ages 10 to 13)—a period thought to span the transition to adolescence. Thus, participants completed self-reports of both pubertal development and empathic ability at age 10 and again at age 13 (see details in the next section). Then, at age 13 (and at least 1 day after the completion of the self-reports), participants underwent fMRI while they believed they were observing another adolescent (who was their same age and gender) being socially excluded by others. To simulate this exclusion, participants observed two rounds of the computer task "Cyberball" (Williams, Cheung, & Choi, 2000; Williams et al., 2002), during which three "players" (who were actually computer programmed) ostensibly played a ball-tossing game together. On their screen, participants could see three icons representing each "player," and they watched as the icons (supposedly controlled by the players) passed the ball back and forth. Prior to the scan, the experimenter explained that the purpose of the study was to examine neural activity during the observation of social interaction. Participants were then told that three individuals their age had volunteered to play the game via the Internet during their scan, and they were given the first names and genders of these "previous participants" (one male, one female, and a third player, who was to be "excluded" by the first two, whose gender matched that of the participant). Participants were instructed to watch the game closely and think about what the players might be thinking or feeling, how they were treating each other, and what strategies they might have for deciding the recipient of each ball toss. In the first round, participants observed all players being included in the game equally (60 throws total). In the second round, one player was excluded by the others after being included for 10 throws and was left out for the rest of the game. Exclusion was used as a proxy for peer rejection based on evidence that isolating peers from social groups is a common way of rejecting peers in adolescence (Coie, Dodge, & Kupersmidt, 1990).

Evidence of Task Validity

Cyberball has elicited feelings of distress and empathy for observed victims of exclusion in prior neuroimaging studies (Beeney et al., 2011; Masten et al., 2010; Masten et al., 2011; Meyer et al., 2012). In addition, careful measures taken in this study support Cyberball's ecological validity.

Evidence from previous literature. First, previously published data (Masten et al., 2010) provide evidence for the effectiveness of Cyberball. After watching the exclusion round of Cyberball, participants reported their state empathy (i.e., how much it hurt to watch the victim being excluded). These reports of state empathy indicated that participants experienced empathic feelings as they watched the exclusion. Next, participants wrote e-mails to the excluded victims, which were later coded for evidence of prosocial behavior (e.g., offering support or comfort). Participants displayed moderate amounts of prosocial behavior, again suggesting that they were concerned for and wanted to help the observed victim (see Masten et al., 2010, for details).

Cover story. In this study, extensive efforts were made to maintain the cover story. Participants were told that the "players" were real kids who had already participated in the study. Participants were even given the chance to volunteer to be a player in a future participant's game. Researchers also pretended to call the players to establish their Internet connections, and once during the scan participants were told to wait while one of the players used the restroom.

Manipulation check. Participants completed a manipulation check to ensure that they noticed the social exclusion during Cyberball. After the scan, they were asked to answer several yes or no questions about whether specific events happened during the game (e.g., "One player was treated unfairly," "All the players participated in the game the same amount"). They were told that this was necessary "because each set of players acts differently." All participants included in this study indicated that one player had been left out during the second round of the game.

Debriefing. After collection of all measures, participants were thoroughly debriefed about the deception in the study. They were also questioned to make sure that they believed that they were observing real people during the scan. All participants reported believing the cover story.

Self-Report Measures

Pubertal development. Participants completed the Pubertal Development Scale (PDS; Petersen, Crockett, Richards, & Boxer, 1988) at ages 10 and 13. On the PDS, participants self-report their visible development of secondary sexual characteristics (e.g., growth spurt, pubic hair, etc.) using a scale from 1 (*no development*) to 4 (*development already completed*) for five items. The PDS has shown good reliability (αs = .68 to .83) and validity (e.g., correlations with physician ratings: *r*s = .61 to .67; correlations with the sexual maturity scale—another self-report puberty measure; see Duke, Litt, & Gross, 1980; Morris & Udry, 1980; *r*s = .72 to .80) in prior samples of early adolescents (see Brooks-Gunn, Warren, Rosso, & Gargiulo, 1987; Petersen et al., 1988). Reliability in the current sample was also good (α = .81). Participants' composite PDS scores at age 13 reflect their concurrent pubertal development. Scores at age 13, after controlling for scores at age 10, reflect changes in participants' pubertal development during the 3 years preceding the scan. To control for scores at age 10, residualized scores for age 13 were calculated, such that the group-level variance in scores at age 13 that was explained by scores at age 10 was removed.

Empathic ability. To obtain indices of distinct aspects of empathy that tapped both mentalizing and affective pain processes, at ages 10 and 13 participants completed three subscales of the Interpersonal Reactivity Index (IRI; Davis, 1983, 1996) measuring: (a) empathic concern—a type of affective empathy (e.g., "When I see someone being taken advantage of, I feel kind of protective toward them"), (b) personal distress—another type of affective empathy (e.g., "I sometimes feel helpless when I am in the middle of a very emotional situation"), and (c) perspective taking—a type of cognitive empathy (e.g., "I try to look at everybody's side of an argument before I make a decision"). Each of these subscales contains seven items that are rated using a scale ranging from 1 (*does not describe me at all*) to 5 (*describes me very well*). The IRI subscales have demonstrated satisfactory internal reliabilities (ranging from .71 to .77) and test–retest reliabilities (ranging from .62 to .71) in prior samples (Davis, 1983), as well as expected associations with related psychological variables, supporting their validity (see Davis, 1983). Satisfactory reliability was also found in the current sample (empathic concern: α = .79; personal distress: α = .74; perspective taking: α = .79). Subscale scores at age 13 reflect 3 dimensions of participants' concurrent empathic ability. Scores at age 13, after controlling for scores at age 10, reflect changes in these dimensions during the 3 years preceding the scan.

fMRI Data Acquisition and Analysis

Data acquisition. Using a Siemens Allegra 3-Tesla MRI scanner, images were collected during functional scans lasting 2 min, 48 s (echo planar T2*-weighted gradient-echo, TR = 2000 ms, TE = 25 ms, flip angle = 90°, matrix size 64 × 64, 36 axial slices, FOV = 20 cm; 3 mm thick, skip 1 mm) for each round of Cyberball, and during a high-resolution structural scan (echo planar T2-weighted spin-echo, TR = 4000 ms, TE = 54 ms, matrix size 128 × 128, FOV = 20 cm, 36 slices, 1.56 mm in-plane resolution, 3 mm thick) to enable functional image registration.

Data analysis. Neuroimaging data were preprocessed and analyzed using SPM5 (Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK), and ROI extraction was performed with the Marsbar toolbox in SPM (http://marsbar.sourceforge.net). Preprocessing included image realignment to correct for head motion, normalization into a standard stereotactic space defined by the Montreal Neurological Institute and the International Consortium for Brain Mapping by averaging 152 brains, and spatial smoothing with an 8 mm Gaussian kernel, full width at half maximum, to increase signal-to-noise ratio.

A block design was used. Each round of Cyberball was modeled as a run with inclusion and exclusion conditions modeled as blocks in the run. Linear contrasts were calculated for each condition comparison for each participant, and used in whole-brain, random-effects analyses.

To examine relations between brain activity during observed exclusion versus inclusion and self-report measures, we used whole-brain and region of interest (ROI) regression analyses. The purpose of the whole-brain analyses was to look at the entire brain volume for any areas in which individuals' brain activity while witnessing peer exclusion related to pubertal development and empathic ability. Specifically, this analysis identified significant clusters of activation representing the particular areas of the brain in which the difference in activity during observed exclusion versus inclusion among our participants, significantly related to their individual differences in concurrent and longitudinal self-reports. In contrast, the purpose of the ROI analyses was to examine the relation between each self-report measure and the average activity during observed exclusion versus inclusion within specific regions identified ahead of time as being meaningful for our task (see ROI definition in the following section). In other words, ROI analyses allowed us to identify whether the average amount of activity displayed by participants, during observed exclusion versus inclusion, in specific a priori–defined brain regions, significantly related to their individual differences in concurrent and longitudinal self-reports.

Thus, overall we examined how differences in neural activity during observed exclusion versus inclusion at age 13, across the whole brain and in a priori–defined ROIs, correlated with: (a) concurrent pubertal development, (b) longitudinal changes in pubertal development from ages 10 to 13, (c) concurrent empathic ability (i.e., empathic concern, personal distress, and perspective taking IRI subscale scores), and (d) longitudinal changes in empathic ability (i.e., changes in empathic concern, personal distress, and perspective taking) from ages 10 to 13.

Whole brain analyses. First, we examined the whole brain to localize areas of activity that were most associated with pubertal development and empathic ability. For each regressor (i.e., self-report measure), we performed one correlational test (the results of which are reported as both t values and r values), which revealed the specific brain areas in which the regressor was significantly associated with the difference in activity between observed exclusion and inclusion. Because standard whole brain analysis software only permits the identification of clusters of activity above a specified threshold, nonsignificant results are not reported as part of our results.

Consistent with Masten et al. (2010), whole brain analyses were thresholded at $p < .005$ for magnitude (uncorrected, minimum cluster size = 10 voxels), for all a priori–defined regions in mentalizing networks (DMPFC, MPFC, TPJ, pSTS, PCC, temporal poles, precuneus), and affective pain networks (dACC, AI, amygdala, subACC; Masten et al., 2009). Other regions were examined at an FDR-corrected threshold ($p < .05$; see Lieberman & Cunningham, 2009). Coordinates are reported in Montreal Neurological Institute (MNI) format.

ROI analyses. Next, we examined adolescents' pubertal development and empathic ability in relation to brain activity during observed exclusion versus inclusion in a set of a priori defined ROIs. These ROIs were functionally defined (using the Marsbar toolbox) as the clusters of activity in the mentalizing network that were found to be significantly more active during the main effect of observed exclusion compared with inclusion in this sample of adolescents (see Masten et al., 2010). Specifically, these ROIs included: one region in the DMPFC (peak voxel [12 44 50]), $t(15) = 4.76$, $p = .0005$, $k = 194$ voxels; two regions in the MPFC (peak voxel [16 70 12]), $t(15) = 3.22$, $p = .005$, $k = 13$ voxels, and (peak voxel [14 46 14]), $t(15) = 3.96$, $p = .001$, $k = 54$ voxels; two regions in the precuneus (peak voxel [10 −66 48]), $t(15) = 4.18$, $p = .0005$, $k = 433$ voxels, and (peak voxel [−12 −62 50]), $t(15) = 3.61$, $p = .005$, $k = 22$ voxels; and one region in the pSTS (peak voxel [58 −44 20]), $t(15) = 3.53$, $p = .005$, $k = 311$ voxels (Masten et al., 2010). Mean parameter estimates for each participant (which model the amplitude of the BOLD response during observed exclusion vs. inclusion) were then extracted and averaged across all voxels in each ROI. Standard software (SPSS 16.0; SPSS, Chicago, IL) was used to conduct regressions to determine whether these parameter estimates were related to each self-reported variable of interest. A standard statistical threshold of $p < .05$ was used for these ROI analyses; based on a priori hypotheses, all tests were one-tailed.

Exploratory mediation analyses. Exploratory mediation analyses were conducted to look at potential directional relations among pubertal development, empathic ability, and neural responses to observed peer rejection. As traditional mediation tests can be biased when used with small samples (Preacher & Hayes, 2008), we employed a bootstrapping method (scripted in SPSS; Preacher & Hayes, 2004) with bias-corrected confidence intervals to more accurately interrogate the significance of these mediational models (Preacher & Hayes, 2004; Shrout & Bolger, 2002). Bootstrap analyses and estimates were based on 2,000 bootstrap samples.

Results

Results for Self-Reported Variables

Descriptive information. As expected, participants' PDS scores reflected significantly greater pubertal development at age 13 (range = 1.20 to 3.40, $M = 2.47$, $SD = .71$) compared to age 10 (range = 1.00 to 1.80, $M = 1.41$, $SD = .28$), $t(15) = 6.32$, $p < .001$. In terms of empathic ability, participants showed significant increases in empathic concern from age 10 (range = 2.00 to 4.43, $M = 3.24$, $SD = .60$) to age 13 (range = 2.29 to 4.43, $M = 3.51$, $SD = .57$), $t(15) = 1.90$, $p < .05$, but their reports of personal distress (age 10: range = 2.00 to 3.00, $M = 2.60$, $SD = .37$; age 13: range = 1.57 to 4.43, $M = 2.69$, $SD = .68$), $t(15) = .42$, *ns*, and perspective taking (age 10: range = 1.57 to 4.50, $M = 3.17$, $SD = .69$; age 13: range = 2.00 to 4.29, $M = 2.90$, $SD = .55$), $t(15) = -1.01$, *ns*, did not significantly change across time.

Gender differences. Given that our sample size was not large enough to meaningfully examine gender differences, gender differences in pubertal development and empathic ability (examined using a series of one-way analyses of variance) are reported here for descriptive purposes only. We do not focus on gender differences in the remainder of the results; however, it is worth noting that controlling for gender did not meaningfully change any of our reported findings.

Boys and girls reported statistically equivalent pubertal development at age 10 (boys: $M = 1.46$, $SD = .28$; girls: $M = 1.38$, $SD = .29$), $F(1, 14) = .31$, *ns*; however, girls reported higher levels by age 13 (boys: $M = 2.09$, $SD = .61$; girls: $M = 2.77$, $SD = .66$), $F(1, 14) = 4.51$, $p = .05$, and greater changes in pubertal development from ages 10 to 13, $F(1, 14) = 6.97$, $p < .01$. In terms of empathic ability, boys reported marginally higher empathic concern than girls at age 10 (boys: $M = 3.49$, $SD = .79$; girls: $M = 3.05$, $SD = .34$), $F(1, 14) = 2.34$, $p = .07$, but girls reported marginally higher empathic concern than boys at age 13 (boys: $M = 3.27$, $SD = .74$; girls: $M = 3.70$, $SD = .33$), $F(1, 14) = 2.51$, $p = .07$, as well as greater changes in empathic concern from ages 10 to 13, $F(1, 14) = 13.47$, $p < .005$. Although girls and boys reported similar personal distress at age 10 (boys: $M = 2.49$, $SD = .26$; girls: $M = 2.69$, $SD = .43$), $F(1, 14) = 1.21$, *ns*, girls reported significantly higher personal distress by age 13 (boys: $M = 2.18$ $SD = .43$; girls: $M = 3.08$ $SD = .57$), $F(1, 14) = 11.80$, $p < .005$, and greater changes in personal distress from ages 10 to 13, $F(1, 14) = 12.58$, $p < .005$. Girls and boys did not differ in their reports of perspective taking at either time point: age 10 (boys: $M = 3.40$, $SD = .71$; girls: $M = 2.98$, $SD = .65$), $F(1, 14) = 1.47$, *ns*, and age 13 (boys: $M = 2.86$, $SD = .77$; girls: $M = 2.94$, $SD = .34$), $F(1, 14) = .09$, *ns*, and there was no gender difference in the degree to which perspective-taking ability changed from ages 10 to 13, $F(1, 14) = .02$, *ns*.

Associations among self-report variables. Intercorrelations of self-report variables are displayed in Table 1. At age 13, pubertal development was positively related to personal distress and marginally related to empathic concern. The relation between pubertal development and perspective taking was positive but not significant. In terms of relations among the dimensions of empathic ability, empathic concern was significantly related to per-

Table 1

Intercorrelations of Self-Reported Pubertal Development and Three Dimensions of Empathic Ability

	1	2	3	4
1. Pubertal development	—	.38†	.42*	.31
2. Empathic concern	.50*	—	.32	.55*
3. Personal distress	.57*	.47**	—	−.02
4. Perspective taking	.12	.28	−.08	—

Note. Values above the diagonal represent correlations among self-report variables measured at age 13. Values below the diagonal represent correlations among increases in self-report variables measured from ages 10 to 13. All correlations presented included 14 *df*.
$^{\dagger}p < .10$. $^{*}p \leq .05$. $^{**}p \leq .01$.

spective taking at age 13 but not personal distress, and perspective taking and personal distress were not related. In terms of the associations among longitudinal changes, increases in puberty from ages 10 to 13 related to increases in empathic concern and personal distress, but not perspective taking. Increases in empathic concern were significantly related to increases in personal distress but not perspective taking, and increases in perspective taking and personal distress were not related.

Neuroimaging Results: Associations With Pubertal Development

First, we examined if (a) adolescents' concurrent pubertal status at age 13 and (b) their changes in pubertal development from ages 10 to 13, were associated with neural activity during observed exclusion versus inclusion at age 13, in regions linked to mentalizing and affective pain processing. We examined correlations across the entire brain, and in a priori–defined ROIs.

Whole brain analyses. Adolescents with greater pubertal development at age 13 showed more activity in regions linked to mentalizing processes, as they watched exclusion versus inclusion. Specifically, they showed greater activity in the bilateral DMPFC (2 58 38; −8 46 28), PCC/precuneus (10 −56 44), bilateral TPJ (48 −50 30; −42 −60 40), and temporal pole (42 20 −32; 52 −2 −34 with anterior spread covering inferior temporal gyrus and temporal pole). Table 2(a) lists details of activations. There were no associations between pubertal development at age 13 and activity during observed exclusion versus inclusion in affective pain network regions and no negative correlations between pubertal developmental at age 13 and brain activity.

Among adolescents who showed greater increases in pubertal development from ages 10 to 13, there was more differential activity during observed exclusion compared to inclusion in two areas of the DMPFC (−10 46 28; −4 60 36; see online Figure S1A) and in the temporal pole (46 −8 −34 with anterior spread covering inferior temporal gyrus and temporal pole). Table 2(b) lists details of activations. There were no associations between changes in pubertal development from ages 10 to 13 and activity during observed exclusion versus inclusion in affective pain network regions, and no negative correlations between increases in pubertal developmental and brain activity.

ROI analyses. Adolescents with greater pubertal development at age 13 showed a marginally greater difference in activity during observed exclusion versus inclusion in one precuneus ROI (10 −66 48), $r(14) = .41$, $p = .056$, in the pSTS ROI, $r(14) = .37$, $p = .082$, and in one MPFC ROI (16 70 12), $r(14) = .36$, $p = .083$. The links between pubertal development at age 13 and activity in the other ROIs were in the expected positive direction but

Table 2

Brain Activity During Observed Exclusion Versus Inclusion Correlated With: (a) Concurrent and (b) Changes in Pubertal Development

	(a) Concurrent puberty						(b) Changes in puberty					
Region	BA		[*x y z*]	*t*(14)	*r*	*p*	BA		[*x y z*]	*t*(14)	*r*	*p*
DMPFC	8	R	[2 58 38]	4.61	.78	<.0005	9	L	[−4 60 36]	3.96	.73	<.001
	9	L	[−8 46 28]	3.70	.70	<.005	9	L	[−10 46 28]	3.52	.69	<.005
PCC	7	R	[10 −56 44]	3.76	.71	<.005						
TPJ	40	R	[48 −50 30]	3.50	.68	<.005						
	40	L	[−42 −60 40]	4.52	.77	<.0005						
Temporal pole	38	R	[42 20 −32]	3.60	.69	<.005						
Temporal pole–ITG	21	R	[52 −2 −34]	4.58	.77	<.0005	21	R	[46 −8 −34]	4.03	.73	<.001

Note. BA = Brodmann's Area; L–R = left–right hemisphere; *x y z* = MNI coordinates; *t* = *t* score of local maxima; *r* = correlations between regressors and brain activity; DMPFC = dorsomedial prefrontal cortex; PCC = posterior cingulate cortex; TPJ = temporal-parietal junction; ITG = inferior temporal gyrus.

not significant. In terms of changes in puberty from ages 10 to 13, there were no associations with brain activity during observed exclusion versus inclusion in any ROIs.

Neuroimaging Results: Associations With Indices of Empathic Ability

Next, we examined if (a) adolescents' concurrent empathic ability (empathic concern, personal distress, and perspective taking IRI subscale scores at age 13) and (b) their changes in this empathic ability from ages 10 to 13, were associated with neural activity during observed exclusion versus inclusion at age 13, in regions involved in mentalizing and affective pain processing. We examined correlations across the entire brain, and in a priori–defined ROIs.

Whole brain analyses. Adolescents with greater empathic concern at age 13 displayed more activity during observed exclusion versus inclusion in regions previously linked with mentalizing—specifically, the bilateral DMPFC (4 54 42; −20 44 38), the MPFC (20 68 12), and the TPJ (−40 −52 40), as well as in a region previously linked with affective pain processing—the AI (44 30 −10). Adolescents with greater personal distress at age 13 also displayed more activity during observed exclusion versus inclusion in two areas of the DMPFC (−2 64 30; −4 54 44), and in the bilateral AI (36 20 −6; −38 14 −10). Adolescents with greater perspective-taking ability at age 13 showed more activity during observed exclusion versus inclusion in two mentalizing regions—the MPFC (2 68 4) and precuneus (4 −70 64), but showed no differential activity in any affective pain-related regions. Table 3(a) lists details of activations. There were no negative correlations between IRI subscale scores at age 13 and brain activity.

In terms of changes in empathic ability from ages 10 to 13, adolescents with greater increases in empathic concern showed more differential activity during observed exclusion versus inclusion in a network of regions involved in mentalizing,

Table 3
Brain Activity During Observed Exclusion Versus Inclusion Correlated With (a) Concurrent and (b) Changes in Empathic Ability

	(a) Concurrent empathy ability						(b) Changes in empathic ability					
Region	BA		[*x y z*]	*t*(14)	*r*	*p*	BA		[*x y z*]	*t*(14)	*r*	*p*
Empathic concern subscale of IRI												
DMPFC	8	R	[4 54 42]	4.70	.78	<.0005	8	R	[10 44 42]	9.01	.92	<.0001
	8	L	[−20 44 38]	4.39	.76	<.0005	8	R	[4 54 42]	5.00	.80	<.0001
MPFC	10	R	[20 68 12]	4.26	.75	<.0005	10	L	[−2 66 2]	4.19	.75	<.0005
							10	L	[−2 66 24]	4.08	.74	<.001
TPJ	40	L	[−40 −52 40]	3.85	.72	<.001	40	R	[54 −52 38]	4.53	.77	<.0005
							40	L	[−44 −50 40]	3.39	.67	<.005
Temporal pole							38	R	[44 12 −36]	3.13	.64	<.005
							38/47	L	[−36 20 −28]	5.02	.80	<.0001
AI		R	[44 30 −10]	4.85	.79	<.0005		R	[44 24 −10]	8.65	.92	<.0001
								L	[−42 14 −14]	4.39	.76	<.0005
SubACC							25/11	L	[−8 30 −12]	3.30	.66	<.005
Personal distress subscale of IRI												
DMPFC	9	L	[−2 64 30]	4.21	.75	<.0005	9	L	[−2 64 30]	4.33	.76	<.0005
	8	L	[−4 54 44]	3.94	.73	<.001	8	L	[−4 54 48]	4.06	.74	<.001
AI		R	[36 20 −6]	3.29	.66	<.005		R	[36 20 −6]	3.29	.66	<.005
		L	[−38 14 −10]	3.39	.67	<.005		L	[−40 12 −12]	3.34	.67	<.005
								L	[−36 32 −14]	3.46	.68	<.005
Perspective taking subscale of IRI												
MPFC	10	R	[2 68 4]	3.82	.71	<.001						
Precuneus	7	R	[4 −70 64]	3.17	.65	<.005	7	R	[2 −68 64]	3.52	.69	<.005
TPJ							40	L	[−40 −52 44]	4.11	.74	<.001
AI								R	[50 6 6]	3.53	.69	<.005

Note. BA = Brodmann's Area; L–R = left–right hemisphere; *x y z* = MNI coordinates; *t* = *t* score of local maxima; *r* = correlations between regressors and brain activity; DMPFC = dorsomedial prefrontal cortex; MPFC = medial prefrontal cortex; TPJ = temporal–parietal junction; AI = anterior insula; subACC = subgenual anterior cingulate cortex.

including: two areas in the DMPFC (10 44 42; 4 54 42; see online Figure S1B), two areas in the MPFC (−2 66 2; −2 66 24), two areas in the temporal poles (44 12 −36; −36 20 −28), and the bilateral TPJ (54 −52 38; −44 −50 40). These adolescents also displayed greater activity in affective pain-related regions, including bilateral AI (44 24 −10; −42 14 −14) and subACC (−8 30 −12). Adolescents with greater increases in personal distress from ages 10 to 13 showed more activity during observed exclusion versus inclusion in regions associated with mentalizing (DMPFC: −2 64 30; −4 54 48; see online Figure S1C) and affective pain processing (AI: 36 20 −6; −40 12 −12; −36 32 −14). Finally, adolescents with greater increases in perspective taking from ages 10 to 13 showed more activity during observed exclusion versus inclusion in regions linked to mentalizing (precuneus: 2 −68 64; TPJ: −40 −52 44) and affective pain processing (AI: 50 6 6). Table 3(b) lists details of activations. There were no negative correlations between changes in IRI scores and brain activity.

ROI analyses. Adolescents with greater empathic concern at age 13 displayed more activity during observed exclusion versus inclusion in the DMPFC, $r(14) = .55$, $p < .05$, and in one of the MPFC (16 70 12), $r(14) = .61$, $p < .01$, ROIs. This association was not significant for the other ROIs. There were no significant associations between brain activity in any of the ROIs and adolescents' personal distress or perspective-taking ability at age 13.

Next, adolescents with greater increases in empathic concern from ages 10 to 13 displayed more activity during observed exclusion versus inclusion in the DMPFC, $r(14) = .66$, $p < .005$; pSTS, $r(14) = .44$, $p < .05$; and one of the MPFC, [16 70 12], $r(14) = .64$, $p < .005$, ROIs. The relations between increases in empathic concern and activity in the other ROIs were in the expected positive direction, but not significant. There were no significant links between ROI activity and increases in adolescents' personal distress or perspective taking from ages 10 to 13.

Results of Exploratory Mediation Analyses

After separately examining how pubertal development and empathic ability each related to neural responses to observed exclusion, we performed mediational tests to explore possible interrelations among these three variables. For these analyses, we focused on empathic concern as our index of empathic ability because it was found to be the most consistently linked to pubertal development and neural activity in the analyses above (i.e., personal distress did not relate to any ROIs, and perspective taking did not relate to any ROIs or pubertal development). We also used estimated neural activity from ROIs (rather than estimates extracted from whole brain analyses) in these mediation tests, so that indices of neural activity would be independent from the measures of pubertal development and empathic ability. Moreover, we focused on ROIs that we identified as being the most likely candidates for mediation based on the above analyses. Specifically, we chose the MPFC (16 70 12) ROI when testing concurrent relations among puberty, empathic concern, and neural activity at age 13, as well as both the MPFC (16 70 12) and pSTS ROIs when testing relations among neural activity at age 13 and longitudinal increases in pubertal development and empathic concern from ages 10 to 13. These ROIs were identified as likely candidates for mediation because they had meaningful positive associations with each of the two other variables included in each model. As the bootstrapping method is specifically designed to permit exploration of significant mediation in small samples that are unlikely to yield p values meeting traditional standards of significance, we used a more liberal threshold of $p < .20$ to identify ROIs that had "meaningful" positive relations with the other variables. The MPFC (16 70 12) was the only ROI showing a meaningful positive relation with both pubertal development ($r = .36$, $p = .08$) and empathic concern ($r = .61$, $p < .01$) at age 13. The MPFC (16 70 12) and pSTS were the only ROIs showing meaningful positive relations with both increases in pubertal development (MPFC: $r = .29$, $p = .14$; pSTS: $r = .36$, $p = .08$) and increases in empathic concern (MPFC: $r = .64$, $p < .005$; pSTS: $r = .44$, $p < .05$) from ages 10 to 13.

First, we examined if a neural mechanism (i.e., increased MPFC or pSTS while observing peer exclusion) might help explain the positive relation between pubertal development and empathic concern. However, none of these models reached significance. Next, we examined whether a cognitive mechanism (i.e., increasing empathic concern) might help explain the positive relation between pubertal development and neural responses to observed peer exclusion. For concurrent relations at age 13, we found no significant mediation. Interestingly, however, we did find that increases in empathic concern from ages 10 to 13 significantly mediated the link between increases in pubertal development from ages 10 to 13 and MPFC activity during observed peer exclusion at age 13 (95% CI

[.02, .81], $p < .05$). Of course, these mediation tests are only preliminary due to the small sample size tested. But, these findings provide initial support for the notion that increasing empathic concern across the transition to adolescence might help explain the positive link between pubertal development across this transition and adolescents' MPFC responses to observed peer exclusion.

Discussion

The findings in this investigation provide new information about how two important indices of adolescent development—pubertal development and empathic ability—relate to emotional and cognitive neural responses while witnessing peer rejection. Our findings show positive links between pubertal development and empathic ability across the adolescent transition and provide preliminary evidence that neural responses to witnessed peer exclusion are associated with both concurrent levels of, and longitudinal changes in, these two developmental indices. Next, we discuss the significance of these findings and their potential implications for understanding adolescents' responses to peer interactions that they witness in their daily lives.

First, correlations among self-report variables indicated both concurrent and longitudinal associations between pubertal development and empathic ability. Specifically, puberty was positively related to empathic concern and personal distress at age 13. Additionally, there were positive relations between changes in pubertal development and changes in both empathic concern and personal distress from ages 10 to 13, which were not evident for perspective taking. These longitudinal findings provide some initial evidence that increases in affective aspects of empathic ability may be most likely to accompany pubertal changes in early adolescence. This is consistent with prior work that has linked pubertal development and social emotion processing (Burnett et al., 2010), social evaluative stress (Gunnar et al., 2009; Sumter et al., 2010), and other types of emotional reactivity (e.g., Silk et al., 2009; Spear, 2009). Of course, it would be premature to dismiss a potential link between pubertal development and perspective taking based on these data, but these findings nevertheless highlight the importance of examining related changes in pubertal development and emotional functioning during the adolescent transition.

In terms of links with neural functioning, we separately examined how pubertal development and empathic ability each related to neural responses to observed peer exclusion. First, we found that adolescents who reported higher levels of physical maturation at age 13 displayed more evidence of cognitive empathy-related processes at the neural level (i.e., more activity in regions involved in mentalizing). This is consistent with the notion that puberty onset is accompanied by a greater sensitivity toward peers, as well as with the common assertion that pubertal development may trigger a host of social cognitive advances during adolescence—evident in both behavior and neural function (see Blakemore et al., 2010; Forbes & Dahl, 2010; Nelson et al., 2005; Steinberg, 2004, 2007). Furthermore, this finding builds on the small body of prior empirical work linking pubertal development, neural activity, and social cognitive processing in adolescence (Forbes et al., 2010; Keulers et al., 2010; Martin et al., 2002; Moore et al., 2012; Silk et al., 2009; Spear, 2009).

Next, using a whole brain analysis we also found that increases in pubertal development from ages 10 to 13 related to heightened activity in the DMPFC and temporal pole when observing a peer being rejected during adolescence. This suggests that not only the concurrent level of pubertal development, but also the degree of change in pubertal development that occurs across the adolescent transition, is positively associated with differential processing in regions linked with mentalizing, when observing a peer being excluded. Perhaps, as adolescents experience pubertal change as a significant life stressor (see Ge et al., 2003), individuals who endure more rapid change may become particularly sensitive to others' responses to stress, based on insights from their own experiences. Of course, conclusive interpretations are not possible at this early stage; however, this finding builds on prior work indicating that the velocity of pubertal change across adolescence is an important predictor of psychosocial outcomes (e.g., Ge et al., 2003; Mendle et al., 2010) and highlights the importance of examining the rate of pubertal change as an unique predictor of adolescent outcomes. We did not find that concurrent, or longitudinal changes in, pubertal development related to activity in affective pain-related regions while observing peer rejection. As this is somewhat inconsistent with our finding that pubertal development is positively related to self-reported empathic concern and personal distress, additional research will be useful for disentangling how pubertal change is associated with particular cognitive and affective aspects of empathic processing.

Next, consistent with our hypotheses regarding empathic ability, we found that adolescents who reported more affective aspects of empathy—specifically, greater empathic concern or personal distress at age 13—displayed more neural evidence of both mentalizing and affective pain processing when observing a peer being excluded. This supports the idea that individuals with a greater tendency to feel concern for others and experience their emotions vicariously may better understand the plight of others and feel more distress when they see them being rejected. In fact, it is possible that the heightened activity in affective pain-related regions among these highly empathic individuals reflects some sharing of the victim's pain or efforts to imagine the victim's distress. In terms of perspective-taking ability, adolescents who reported higher perspective-taking ability at age 13 also displayed more neural evidence of mentalizing, but not affective pain processing. This is consistent with our hypothesis and suggests that individuals who are better at perspective taking may be more likely to spontaneously think about the thoughts and feelings of others that they see engaged in social interactions.

It is worth noting that adolescents' reports on these three dimensions of empathy related to neural activity in different ways. In other words, the more affective dimensions—empathic concern and personal distress—were associated with neural activity in regions linked to both mentalizing and affective pain processing, whereas perspective taking was only associated with activity in regions linked with mentalizing. Thus, these findings build on our previous work, in which we focused on a unidimensional measure of empathy and only found associations with activity in the mentalizing network and not the network of regions involved in affective pain processing (Masten et al., 2010). Of course, these patterns warrant further examination in future research; however, these findings highlight the utility of using multidimensional measures of empathy when examining empathy for social interactions—particularly at the neural level. Additionally, these findings raise another intriguing possibility—that perspective taking might be more strictly cognitive in nature without necessarily having an emotional element, whereas emotional aspects of empathy might require some mentalizing (as evidenced by the presence of activation in regions linked to both mentalizing and affective pain processing) in order to understand a peer's plight enough to display an appropriate emotional response. In fact, the cognitive ability to understand others' emotions might be particularly important early in the empathic process and help facilitate more emotional aspects of empathy (i.e., concern, sharing of others' emotions) among those who show heightened affective empathic tendencies (see Masten et al., 2011, for a discussion of this possibility).

In terms of longitudinal changes in empathic ability across the adolescent transition, increases in all three dimensions of empathic ability from ages 10 to 13 were related to greater activity in neural regions linked to both mentalizing and affective pain activity. Thus, although some research shows that different aspects of empathic ability develop via different trajectories (i.e., perspective taking typically increases, while personal distress decreases; Eisenberg, Cumberland, Guthrie, Murphy, & Shepard, 2005), these data suggest that increases on any of these dimensions at the individual level is linked with greater empathy-related neural activity in both mentalizing and affective pain-related regions. One possibility is that the social changes that occur during the transition to adolescence—such as increases in peer interactions—necessitate improvements in overall empathic ability. In other words, there may be an increased need for empathy and mentalizing in order for adolescents to successfully navigate their changing social climate (see Eisenberg & Morris, 2004; Eisenberg et al., 2009). In this case, greater neural sensitivity during observed peer rejection might characterize those individuals who have been most successful in terms of improving their empathic ability. Another way of conceptualizing this possibility is that more rapid increases in empathic ability and greater neural sensitivity to observed peer rejection are both characteristics of individuals who are more socially sensitive and responsive to others at a trait level. In other words, upon entering adolescence, these individuals may be more sensitive to their changing surroundings and adjust their empathic tendencies accordingly, and also be more impacted by peer interactions that they witness. Finally, it is worth noting that the correlations between adolescents' neural responses to observed peer rejection and their longitudinal changes in empathic ability were somewhat distinct from the correlations found with concurrent levels of empathic ability. This highlights the utility of measuring developmental variables of interest at multiple time points, particularly when examining brain–behavior relations within developing populations.

Finally, to gain preliminary insights into the possible direction and causality of our effects, we performed exploratory tests of possible mediation pathways via which pubertal development, empathic ability, and neural responses to observed peer exclusion might be interrelated. We found initial evidence that increases in empathic concern that accompany pubertal change from ages 10 to 13 may help explain the heightened neural sensitivity that adolescents display toward their peers. Notably, tests for mediation among these variables measured concurrently at age 13 did not yield significant results, suggesting that examining interrelated patterns of longitudinal change as youth transition to adolescence may be important for understanding the salience of observed peer interactions later in the adolescent years. Along with prior work showing the importance of physical maturation and increasing empathic ability for promoting interest in peers and social acceptance (Eisenberg & Morris, 2004; Eisenberg et al., 2009; Forbes & Dahl, 2010), these findings highlight the need for additional longitudinal investigations of potential contributors to peer salience in adolescence.

Future Directions

Future studies with larger samples will be able to more thoroughly interrogate these relations and determine the order and timing of these effects. For example, it would be interesting to conclusively test whether pubertal maturation leads to improvements in empathic ability, which in turn influence neural function, as our initial findings suggest. Or, will future work uncover evidence in favor of the alternate possibility that changes in neural function precipitated by pubertal onset enable adolescents to better understand others' emotions? Multimethod research with large samples will facilitate continued exploration of these questions. Eventually, uncovering the directionality and temporal order of these interrelated processes will broaden understanding of how social, cognitive, and physical changes unfold in adolescence.

In addition to pursuing these questions, future work will also be useful in addressing the limitations of this study. First, as mentioned previously, our sample size was relatively small and did not permit meaningful exploration of gender or other individual differences (e.g., related to race or ethnicity, socioeconomic status, etc.). Males' and females' social cognitive processing in adolescence may differ in meaningful ways due to their differential trajectories of pubertal onset and development (e.g., girls typically mature earlier than boys). Thus, it will be useful to examine gender and other individual differences, in future studies with larger samples.

Next, it would also be useful for future research to examine other measures of longitudinal biological change and physical maturation, for example, additional indices of puberty such as Tanner Staging or hormonal assays, or changes in brain structure at this age. These additional measures would provide more precise, multidimensional indices of the participants' level of physical and brain maturation, which could reveal additional interesting relations among various measures of physical, social, cognitive, and neural development that were not detectable in this study. Additionally, controlling for potential behavioral correlates of puberty (e.g., risk taking, sensation seeking, changes in self-image) or social contextual factors that are salient at this age (e.g., transitioning to middle school, time spent with peers vs. parents, frequency of firsthand and witnessed peer rejection) would also be useful, because we cannot be sure that pubertal maturation, rather than a related change in behavior or context, lead to our reported findings. Finally, this study used a single task to simulate the experience of observing peer exclusion. Using additional tasks (e.g., proxies for peer rejection such as rejection film clips or "mean" social networking pages, as well as other types of social interaction tasks) in future studies could help determine if the patterns observed here are specific to observing peer exclusion, or if they characterize other social cognitive processes as well.

Conclusion

Overall, these findings provide initial insights into adolescents' responses when they witness peer rejection and suggest that physical and cognitive emotional development across the transition to adolescence relate to these responses at the neural level. This study builds on the limited prior research that has examined adolescents' neural functioning in the context of pubertal development, and in relation to longitudinal indices of development (Pfeifer & Blakemore, 2012). We hope these findings pave the way for new longitudinal research examining the complex interplay of peer relationships and brain function in adolescence.

References

Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and mental health problems:

"Much ado about nothing"? *Psychological Medicine, 40,* 717–729.

Baron-Cohen, S. (2003). *The essential difference.* New York: Basic Books.

Beeney, J. E., Franklin, R. G., Levy, K. N., & Adams, R. B. (2011). I feel your pain: Emotional closeness modulates neural responses to empathically experienced rejection. *Social Neuroscience, 6,* 369–376.

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9,* 267–277.

Blakemore, S.-J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping, 31,* 926–933.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage, 25,* 312–319.

Bramen, J. E., Hranilovich, J. A., Dahl, R. E., Forbes, E. E., Chen, J., Toga, A. W., et al. (2011). Puberty influences medial temporal lobe and cortical gray matter maturation differently in boys than girls matched for sexual maturity. *Cerebral Cortex, 21,* 636–646.

Brendgen, M., & Vitaro, F. (2008). Peer rejection and physical health problems in early adolescence. *Journal of Developmental and Behavioral Pediatrics, 29,* 183–190.

Brooks-Gunn, J., Warren, M. P., Rosso, J., & Gargiulo, J. (1987). Validity of self-report measures of girls' pubertal status. *Child Development, 58,* 829–841.

Brown, B. B. (2004). Adolescents' relationships with peers. In R. Lerner, & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 363–394). Hoboken, NJ: Wiley.

Burnett, S., Thompson, S., Bird, G., & Blakemore, S.-J. (2010). Pubertal development of the understanding of social emotions: Implications for education. *Learning and Individual Differences, 21,* 681–689.

Coie, J. D., Dodge, K. A., & Kupersmidt, J. B. (1990). Peer group behavior and social status. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood: Cambridge studies in social and emotional development* (pp. 17–59). New York, NY: Cambridge University Press.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44,* 113–126.

Davis, M. H. (1996). *Empathy—A social psychological approach.* Boulder, CO: Westview.

Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience, 32,* 257–267.

Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology, 20,* 1053–1080.

Decety, J., Michalska, K. J., & Akitsuki, Y. (2008). Who caused the pain? An fMRI investigation of empathy and intentionality in children. *Neuropsychologia, 46,* 2607–2614.

Denham, S. A., Renwick-DeBardi, S., & Hewes, S. (1994). Emotional communication between mothers and preschoolers: relations with emotional competence. *Merrill-Palmer Quarterly, 40,* 488–508.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Science, 10,* 435–441.

Duke, P. M., Litt, I. F., & Gross, R. T. (1980). Adolescents' self-assessment of sexual maturation. *Pediatrics, 66,* 918–920.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Reviews of Psychology, 51,* 665–697.

Eisenberg, N., Cumberland, A., Guthrie, I. K., Murphy, B. C., & Shepard, S. A. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *Journal of Research on Adolescence, 15,* 235–260.

Eisenberg, N., Miller, P. A., Shell, R., McNally, S., & Shea, C. (1991). Prosocial development in adolescence: A longitudinal study. *Developmental Psychology, 27,* 849–857.

Eisenberg, N., & Morris, A. S. (2004). Moral cognitions and prosocial responding in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 155–188). Hoboken, NJ: Wiley.

Eisenberg, N., Morris, A. S., McDonald, B., & Spinrad, R. L. (2009). Moral cognitions and prosocial responding in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Vol. 1. Individual bases of adolescent development* (pp. 229–265). Hoboken, NJ: Wiley.

Eisenberg, N., Spinrad, T. L., & Sadovsky, A. (2006). Empathy-related responding in children. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 517–549). Mahwah, NJ: Erlbaum.

Forbes, E. E., & Dahl, R. E. (2010). Pubertal development and behavior: Hormonal activation of social and motivational tendencies. *Brain and Cognition, 72,* 66–72.

Forbes, E. E., Ryan, N. D., Phillips, M. L., Manuck, S. B., Worthman, C. M., Moyles, D. L., et al. (2010). Healthy adolescents' neural response to reward: Associations with puberty, positive affect, and depressive symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry, 49,* 162–172.

Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science, 286,* 1692–1695.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50,* 531–534.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of Biological Sciences, 358,* 459–473.

Frith, U., Leslie, A. M., & Morton, J. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences, 14,* 433–438.

Ge, X., Kim, I. J., Brody, G. H., Conger, R. D., Simon, R. L., Gibbons, F. X., et al. (2003). It's about timing and change: Pubertal timing effects on symptoms of depression among African American youths. *Developmental Psychology, 39,* 430–439.

Gunnar, M., Wewerka, S., Frenn, K., Long, J., & Griggs, C. (2009). Developmental changes in hypothalamus-pituitary-adrenal activity over the transition to adolescence: Normative changes and associations with puberty. *Development and Psychopathology, 21,* 69–85.

Isaacs, J., Hodges, E. V. E., & Salmivalli, C. (2008). Long-term consequences of victimization by peers: A follow-up from adolescence to young adulthood. *European Journal of Developmental Science*, *2*, 387–397.

Jackson, P. L., Bruney, E., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, *24*, 771–779.

Janes, L. M., & Olson, J. M. (2000). Jeer pressures: The behavioral effects of observing ridicule of others. *Personality and Social Psychology Bulletin*, *26*, 474–485.

Keulers, H. H., Evers, E. A. T., Stiers, P., & Jolles, J. (2010). Age, sex, and pubertal phase influence mentalizing about emotions and actions in adolescents. *Developmental Neuropsychology*, *35*, 555–569.

Lev-Wiesel, R., Nuttman-Shwartz, O., & Sternberg, R. (2006). Peer rejection during adolescence: Psychological long-term effects—A brief report. *Journal of Loss & Trauma*, *11*, 131–142.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*, 423-428.

Martin, C. A., Kelly, T. H., Rayens, M. K., Brogli, B. R., Brenzel, A., Smith, W. J., et al. (2002). Sensation seeking, puberty, nicotine, alcohol and marijuana use in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 1495–1502.

Masten, C. L., & Eisenberger, N. I. (2009). Exploring the experience of social rejection in adults and adolescents: A social cognitive neuroscience perspective. In M. Harris (Ed.), *Bullying, rejection, and peer victimization: A social cognitive neuroscience perspective* (pp. 53–78). New York: Springer.

Masten, C. L., Eisenberger, N. I., Borofsky, L. A., Pfeifer, J. H., McNealy, K., Mazziotta, J., et al. (2009). Neural correlates of social exclusion during adolescence: Understanding the distress of peer rejection. *Social Cognitive and Affective Neuroscience*, *4*, 143–157.

Masten, C. L., Eisenberger, N. I., Pfeifer, J. H., & Dapretto, M. (2010). Witnessing peer rejection during adolescence: Neural correlates of empathy for experiences of social exclusion. *Social Neuroscience*, *5*, 496–507.

Masten, C. L., Eisenberger, N. I., Pfeifer, J. H., & Dapretto, M. (2012). *Neural responses to witnessing peer rejection after being socially excluded: fMRI as a window into adolescents' emotional processing*. Manuscript submitted for publication.

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *NeuroImage*, *55*, 381–388.

Mendle, J., Harden, K. P., Brooks-Gunn, J., & Graber, J. A. (2010). Development's tortoise and hare: Pubertal timing, pubertal tempo, and depressive symptoms in boys and girls. *Developmental Psychology*, *46*, 1341–1353.

Meyer, M. L., Masten, C. L., Ma, Y., Wang, C., Shi, Z., Eisenberger, N. I., et al. (2012). Interpersonal closeness modulates empathic neural responses to social pain. *Social Cognitive and Affective Neuroscience*. Advance online publication. doi:10.1093/scan/nss019.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*, 1306–1315.

Moore, W. E., Pfeifer, J. H., Masten, C. L., Mazziotta, J. C., Iacoboni, M., & Dapretto, M. (2012). Facing puberty: Associations between pubertal development and neural responses to affective facial displays. *Social Cognitive and Affective Neuroscience*, *7*, 35–43.

Morris, N., & Udry, J. R. (1980). Validation of a self-administered instrument to assess stage of adolescent development. *Journal of Youth and Adolescence*, *9*, 271–280.

Morrison, I., Lloyd, D., Di Pellegrino, G., & Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: Is empathy a multisensory issue? *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 270–278.

Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, *35*, 163–174.

Nishina, A., & Juvonen, J. (2005). Daily reports of witnessing and experiencing peer harassment in middle school. *Child Development*, *76*, 435–450.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*, 117–133.

Pfeifer, J. H., & Blakemore, S.-J. (2012). Adolescent social cognitive and affective neuroscience: Past, present, and future. *Social Cognitive and Affective Neuroscience*, *7*, 1–10.

Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring others' emotions related to empathy and interpersonal competence in children. *NeuroImage*, *39*, 2076–2085.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers*, *36*, 717–731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavioral Research Methods*, *40*, 879–891.

Prinstein, M. J., Sheah, C. S., & Guyer, A. E. (2005). Peer victimization, cue interpretation, and internalizing symptoms: Preliminary concurrent and longitudinal findings for children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 11–24.

Rivers, I., Poteat, P. V., Noret, N., & Ashurst, N. (2009). Observing bullying at school: The mental health implications of witness status. *School Psychology Quarterly*, *24*, 211–223.

Saxe, R. (2006). Why and how to study theory of mind with fMRI. *Brain Research*, *1079*, 57–65.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain: A Journal of Neurology*, *132*, 617–627.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445.

Silk, J. S., Siegle, G. J., Whalen, D. J., Ostapenko, L. J., Ladouceur, C. D., & Dahl, R. E. (2009). Pubertal changes in emotional information processing: Pupillary, behavioral, and subjective evidence during emotional word identification. *Development and Psychopathology*, *21*, 7–26.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, *30*, 855–863.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*, 81–96.

Singer, T., Seymour, B., O'Doherty, J. P., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–1162.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466–469.

Spear, L. (2009). Heightened stress reactivity and emotional reactivity during pubertal maturation: Implications for psychopathology. *Development and Psychopathology*, *21*, 87–97.

Steinberg, L. (2004). Risk-taking in adolescence: What changes, and why? *Annals of the New York Academy of Sciences*, *1021*, 51–58.

Steinberg, L. (2007). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, *28*, 78–106.

Sumter, S. R., Bokhorst, C. L., Miers, A. C., Van Pelt, J., & Westenberg, P. M. (2010). Age and puberty differences in stress responses during a public speaking task: Do adolescents grow more sensitive to social evaluation? *Psychoneuroendocrinology*, *35*, 1510–1516.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion*, *1*, 70–83.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.

Williams, K. D., Cheung, C. K., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology*, *79*, 748.

Williams, K. D., Govan, C. L., Croker, V., Tynan, D., Cruickshank, M., & Lam, A. (2002). Investigations into differences between social- and cyberostracism. *Group Dynamics: Theory Research, and Practice*, *6*, 65–77.

Zahn-Waxler, C., Cole, P. M., Welsh, J. D., & Fox, N. A. (1995). Psychophysiological correlates of empathy and prosocial behaviors in preschool children with behavior problems. *Development and Psychopathology*, *7*, 27–48.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*, 126–136.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Activity During Observed Exclusion Versus Inclusion in the DMPFC, Found Through Whole Brain Regression Analyses to Be Positively Related to Longitudinal Changes in: (a) Pubertal Development, (b) Empathic Concern, and (c) Personal Distress, Across the Transition to Adolescence.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1355–1372

Developmental Trajectories of Acculturation in Hispanic Adolescents: Associations With Family Functioning and Adolescent Risk Behavior

Seth J. Schwartz, Sabrina Des Rosiers, and Shi Huang
University of Miami

Byron L. Zamboanga
Smith College

Jennifer B. Unger
University of Southern California

George P. Knight
Arizona State University

Hilda Pantin and José Szapocznik
University of Miami

This study examined longitudinal acculturation patterns, and their associations with family functioning and adolescent risk behaviors, in Hispanic immigrant families. A sample of 266 Hispanic adolescents ($M_{\text{age}} = 13.4$) and their primary parents completed measures of acculturation, family functioning, and adolescent conduct problems, substance use, and sexual behavior at five timepoints. Mixture models yielded three trajectory classes apiece for adolescent and parent acculturation. Assimilated adolescents reported the poorest family functioning, but adolescent assimilation negatively predicted adolescent cigarette smoking, sexual activity, and unprotected sex indirectly through family functioning. Follow-up analyses indicated that discrepancies between adolescent and parent family functioning reports predicted these adolescent outcomes. Results are discussed regarding acculturation trajectories, adolescent risk behavior, and the mediating role of family functioning.

Rates of international migration have reached unprecedented levels around the world. In many cases, individuals are migrating from developing, collectivist-based countries to postindustrial, individualistic Western countries (Steiner, 2009). This surge in migration has prompted increased attention to acculturation—the ways in which individuals culturally adapt following their move to a new country. Acculturation represents changes in one's cultural orientation that occur as a result of contact with culturally dissimilar people, places, and customs (Berry, 1980). Most studies of acculturation focus on immigrants and on children of immigrants, who have been directly exposed to both their heritage cultures and the culture of the receiving country or region (Portes & Rumbaut, 2001, 2006).

Acculturation has been commonly studied in terms of language use and other cultural practices, such as choice of friends, food and music preferences, and ways of celebrating holidays (Kang, 2006; Unger et al., 2002). Most recent views of acculturation have adopted a bidimensional approach—where heritage culture and receiving-culture orientations are considered separately (e.g., Ryder, Alden, & Paulhus, 2000; Szapocznik, Kurtines, & Fernandez, 1980) and where individuals can endorse both the heritage and receiving cultures. For example, in the case of Hispanic immigrants in the United States—one can "act American" at work or school and "act Hispanic" at home, or can combine the Hispanic and U.S. cultures into a

---

Sabrina Des Rosiers is now at Barry University, Miami Shores, Florida.

This research was supported by Grants DA19409 and DA25694 from the National Institute on Drug Abuse, and by Grant MH63042 from the National Institute on Mental Health.

Correspondence concerning this article should be addressed to Seth J. Schwartz, Department of Epidemiology and Public Health, Leonard M. Miller School of Medicine, University of Miami, 1425 N.W. 10th Avenue, Suite 321, Miami, FL 33136. Electronic mail may be sent to SSchwartz@med.miami.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0018
DOI: 10.1111/cdev.12047

Hispanic-American way of living (e.g., Benet-Martínez & Haritatos, 2005).

Until the beginning of the 21st century, a major limitation of the acculturation literature was a reliance on cross-sectional studies (Fuligni, 2001) where acculturation was treated more as an individual-difference variable than as a trajectory of adaptation over time. In the decade since Fuligni's (2001) call for longitudinal work, a number of studies (e.g., Berkel et al., 2010; Knight, Vargas-Chanes, et al., 2009; Matsunaga, Hecht, Elek, & Ndiaye, 2010; Unger, Ritt-Olson, Wagner, Soto, & Baezconde-Garbanati, 2009) have examined acculturation longitudinally. Other studies have examined longitudinal trajectories of related cultural processes, such as ethnic identity (e.g., Pahl & Way, 2006). These studies have advanced the acculturation literature considerably and have been consistent with Berry's (1980) definition of acculturation as a longitudinal trajectory of cultural adaptation. Moreover, a common theme among these studies has been the identification of multiple trajectories of cultural adaptation. For example, Knight, Vargas-Chanes, et al. (2009), in their 3-year study of Mexican Americans, found two trajectory groups (Mexican and dual cultural) for Hispanic and American cultural practices and three trajectory groups (bilingual, primarily English, and monolingual English) for Spanish and English-language use. Similarly, Matsunaga et al.'s (2010) study of Mexican Americans found movement in multiple directions during the course of adolescence—with some youth shifting toward a strong heritage-cultural orientation and others shifting away from heritage-cultural orientations. Acculturation is clearly not a singular process that proceeds in the same fashion across all individuals.

The majority of longitudinal acculturation studies thus far have been primarily descriptive, in that the objective was to map the course of acculturative change over time. This is not surprising, given that longitudinal acculturation research has emerged largely within the past decade. No published studies have yet linked different trajectories of acculturation to family relationships or to adolescent outcomes. This is an important next step for the field, given evidence that acculturation may be related to family functioning (e.g., Szapocznik & Kurtines, 1993; Unger, Ritt-Olson, Wagner, et al., 2009) and to adolescent behavior problems and risk taking (e.g., Allen et al., 2008; Dinh, Roosa, Tein, & Lopez, 2002; Zamboanga, Schwartz, Jarvis, & Van Tyne, 2009). The heterogeneity in acculturative trajectories found by Knight, Vargas-Chanes, et al. (2009), Matsunaga et al. (2010), and others opens up the possibility of identifying how different acculturative pathways might be differentially associated with family relationships and with adolescent developmental outcomes.

Acculturation and the Family

Although acculturation is often studied as an individual-level construct, in many cases it is also a family-level phenomenon (Crockett & Zamboanga, 2009; Santisteban & Mitrani, 2003), especially for children or adolescents who are raised in immigrant-headed homes. Parents' efforts to preserve the heritage culture are associated with children's continued endorsement and maintenance of aspects of that culture through adolescence (Umaña-Taylor, Bhanot, & Shin, 2006) and into early adulthood (Schwartz & Zamboanga, 2008). On the other hand, immigrant parents may experience trouble handling adolescents who have drifted away from the heritage culture (e.g., Smokowski & Bacallao, 2011). Indeed, parents may interpret individualistic, receiving-culture-oriented adolescent behavior (e.g., spending more time with friends) as disrespectful and as an affront to the collectivist-oriented family (Portes & Rumbaut, 2001).

Research suggests that many adults settling in ethnically concentrated areas may not acculturate much at all (Schwartz, Pantin, Sullivan, Prado, & Szapocznik, 2006). Children and adolescents, who attend school in the receiving society's language and are exposed to receiving-culture media, dress styles, and other influences would be expected to orient themselves more toward the receiving society; at the same time, remaining oriented toward the heritage culture is important in maintaining positive family bonds with traditionally oriented parents (e.g., Unger, Ritt-Olson, Soto, & Baezconde Garbanati, 2009). Biculturalism, where the adolescent endorses the practices of both the heritage and receiving cultures, is often the most adaptive approach to acculturation (e.g., Coatsworth, Maldonado-Molina, Pantin, & Szapocznik, 2005).

Because acculturation is a family-level phenomenon, including family processes in studies of acculturation and risky behavior is important (Unger, Ritt-Olson, Soto et al., 2009; Unger, Ritt-Olson, Wagner, et al., 2009). Aside from the direct effects of acculturation on family functioning, several studies have also examined whether family processes mediate the effects of acculturation on adolescent outcomes. Theory (e.g., Szapocznik & Kurtines, 1993) and research (e.g., Gonzales, Deardorff, Formoso, Barr, & Barrera, 2006; Unger, Ritt-Olson, Soto, et al., 2009; Unger, Ritt-Olson, Wagner, et al., 2009) suggest that within Hispanic families, adolescent acculturation toward American cultural practices is associated with compromised family relationships (e.g., communication, involvement, cohesion), which in turn are predictive of problematic adolescent behaviors such as conduct problems, drug and alcohol use, and unsafe sexual behavior. In some cases, adolescents' retention of Hispanic cultural practices has been associated with more favorable family processes (Smokowski, Rose, & Bacallao, 2008), which may in turn be protective against problematic adolescent behavior. Moreover, only a few studies have used independent parent and adolescent reports of acculturation and family functioning (e.g., Gonzales et al., 2006; Smokowski et al., 2008)—and this is an important methodological advance in the literature.

It is also important to examine the ways in which acculturation *trajectories*—rather than just prior *levels* of acculturation—are predictive of family functioning. In particular, Cicchetti and Rogosch (2002) have discussed the concept of equifinality, where one can arrive at the same endpoint via a number of different trajectories. For example, biculturalism may emerge when a recent immigrant becomes oriented toward American culture in addition to retaining her or his cultural heritage, or when a U.S.-born individual becomes reinvested in her or his cultural heritage in addition to being oriented toward the United States (Schwartz, Unger, Zamboanga, & Szapocznik, 2010). Although these trajectories appear to lead to the same place (i.e., biculturalism), different trajectories may be associated with quite different effects on the family system. For example, adolescents who reorient themselves toward Hispanic culture may engender increased closeness to family members (Szapocznik et al., 1989), whereas adolescents who gravitate toward American culture may create tension within the family (Santisteban & Mitrani, 2003). As a result, it is important to examine the over-time trajectories of acculturation in relation to family functioning in Hispanic adolescents and families.

Acculturation and Health Risk Behavior

A number of studies, most of which have been cross-sectional, have examined direct associations between acculturation and risk behaviors such as substance use and unsafe sexual behavior. In general, these studies have utilized unidimensional models of acculturation where Hispanic and American cultural practices are cast as opposite ends of a single continuum (e.g., Allen et al., 2008; Ramirez et al., 2004). Several studies have also suggested that increased "acculturation" predisposes Hispanic adolescents toward early sexual behavior (Guilamo-Ramos, Jaccard, Peña, & Goldberg, 2005), lower likelihood of condom use (Kepka, Coronado, Rodriguez, & Thompson, 2010), and greater number of lifetime sexual partners (Lee & Hahm, 2010). However, given the reliance on unidimensional models of acculturation, the extant literature leaves open the question of how longitudinal trajectories of more construct-valid indices of acculturation might be associated with sexual behavior and risk taking among Hispanic adolescents.

Although unidimensional models of acculturation do not reflect the lived experiences of most immigrants, longitudinal studies using language use as a proxy for acculturation may support the hypothesis that family processes would mediate the effects of acculturation on adolescent problem and risk outcomes. For example, McQueen, Getz, and Bray (2003) found that Mexican American adolescents' feelings of separation from family members mediated the effects of English-language use on deviant behavior and on cigarette and alcohol use. Mogro-Wilson (2008) found that among Hispanic adolescents, use of English was associated with lower parental control, which in turn predicted adolescent alcohol use. Gil, Wagner, and Vega (2000), using a sample of Hispanic boys in Miami, found that English-language use was linked with lower levels of respect and regard for parents—which in turn was predictive of adolescent alcohol use.

This Study

This study contributes to the growing literature on longitudinal patterns of acculturation by: (a) examining trajectories of acculturation (both American-culture acquisition and Hispanic-culture retention) in adolescents and parents, and then, (b) linking these trajectories to adolescent and parent reports of family functioning and to adolescent reports of risky behavior—conduct problems, substance use, and sexual activity (including unprotected sex). Conduct problems are risky because they may serve as precursors to, or occur along with, drug and alcohol use (Park & Romer, 2010). Sexual activity prior to age 16 years is also risky because it is associated with unprotected sex, greater numbers of sexual partners, and sex under the influence of alcohol or drugs later on (Dillon et al., 2010).

Although data on all variables were collected at each of the five study timepoints, we utilized data on each variable only at certain timepoints so as to specify the temporal order of our study variables (i.e., acculturation trajectories → family functioning → behavior problems). This ordering was based on Szapocznik and Kurtines's (1993) theoretical perspective, which posits that acculturation can exert a causal effect on family processes, which then affect adolescent behavioral outcomes (see Figure 1). Given this theoretical grounding, and so that we could test for longitudinal mediation, we operationalized acculturation as a trajectory across four timepoints spanning 2 years (cf. Matsunaga et al., 2010), and we modeled family functioning reports (from both parents and adolescents) at the fourth study timepoint as a function of this acculturation trajectory. We controlled for family functioning at the third timepoint so that only the unique longitudinal effects of acculturative trajectories on family functioning would be modeled. Similarly, adolescent outcomes were modeled at the fifth (last) timepoint and were controlled for at the fourth timepoint.

Following Knight, Vargas-Chanes, et al. (2009), we used latent class growth analysis (Muthén & Muthén, 2000; Nagin, 2005) to extract classes of adolescents and of parents, where each group would be characterized by similar baseline levels and change trajectories in acculturation. Acculturation trajectories were allowed to predict adolescent conduct problems, substance use, and sexual risk taking both (a) directly and (b) indirectly through family functioning (cf. Gonzales et al., 2006). In line with bidimensional models of acculturation, we created trajectory classes for adolescents and for parents using both Americanism and Hispanicism (cf. Knight, Vargas-Chanes, et al., 2009), and we then examined how these acculturation trajectories were related to family functioning and to adolescent outcomes. This study was conducted in Miami, which serves as a Hispanic ethnic enclave for adult immigrants and as a highly bicultural context for children, adolescents, and young adults (Stepick & Stepick, 2002).

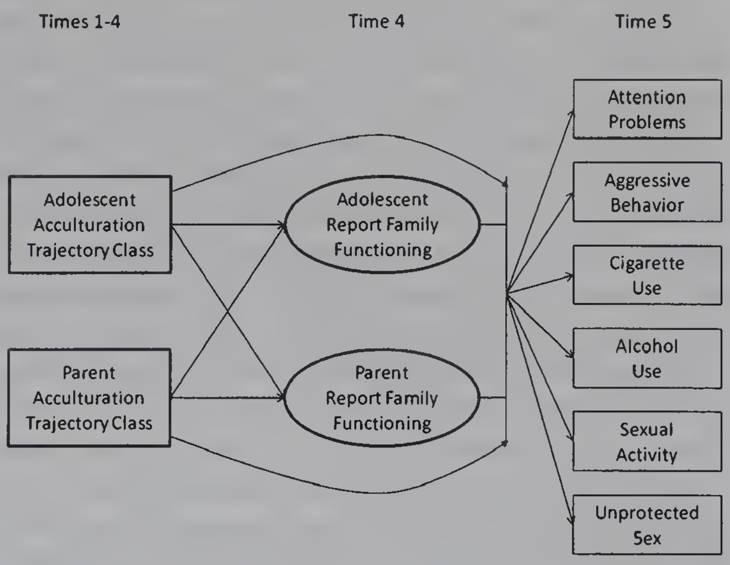

Times 1-4
Time 4
Time 5
Adolescent Acculturation Trajectory Class
Parent Acculturation Trajectory Class
Adolescent Report Family Functioning
Parent Report Family Functioning
Attention Problems
Aggressive Behavior
Cigarette Use
Alcohol Use
Sexual Activity
Unprotected Sex

Figure 1. Structural model.

Hypotheses. Consistent with Berry's (1980) model of acculturation, we anticipated that for adolescents, trajectory classes would emerge representing assimilation (high Americanism combined with low Hispanicism), separation (low Americanism with high Hispanicism), and biculturalism (high Americanism with high Hispanicism). We did not expect a cluster representing marginalization (low Americanism with low Hispanicism), because it is unlikely that individuals living in a bicultural context would fail to endorse either American or Hispanic behaviors over time (Rudmin, 2003). For parents, given that most Hispanic parents who settle in ethnic enclaves retain their cultural heritage and do not endorse American behaviors strongly (Portes & Rumbaut, 2006; Schwartz et al., 2006), we anticipated that the classes extracted would be characterized by high Hispanicism and would differ only in the extent of Americanism endorsed.

Given the associations between low levels of adolescent Hispanicism and family disengagement or conflict (Gil et al., 2000; McQueen et al., 2003), we further hypothesized that adolescent acculturation trajectory classes characterized by high or increasing Hispanicism scores would predict more favorable indices of family functioning (e.g., high parent–adolescent involvement, positive parenting, and communication). On the other hand, an assimilated adolescent trajectory, characterized by low Hispanicism scores over time, would be expected to predict poor family functioning. An assimilated adolescent trajectory would also be expected to predict delinquent and risk-taking behaviors indirectly through family functioning (cf. McQueen et al., 2003). For parents, trajectories characterized by higher levels of Americanism were hypothesized to positively predict delinquent and risk-taking adolescent behaviors, likely because Americanized parents may be more permissive (Dumka, Gonzales, Bonds, & Millsap, 2009).

Method

Participants and Procedures

Data for this study were taken from a randomized clinical trial of Familias Unidas (Grant MH63042; Pantin et al., 2003; Pantin et al., 2009; Prado et al., 2007), a family-based substance use and HIV prevention program for Hispanic adolescents and their parents. Although the data set was taken from an intervention study, there is evidence that intervention data sets can be used for longitudinal analyses as long as: (a) the study model is found to fit equivalently across conditions, and (b) at least one of the conditions does not target any of the study variables (Choi, Harachi, Gillmore, & Catalano, 2005; Matsunaga et al., 2010). Acculturation was not directly targeted by any of the three conditions, family functioning conduct problems and substance use were directly targeted by the first condition, and sexual behavior was directly targeted by the first and second conditions. The third condition did not directly target any of the study variables.

One parent from each family participated in the study with the adolescent. Recruitment took place during April through June of the adolescents' seventh-grade year. During recruitment, on several occasions, study staff visited all seventh-grade homerooms in the three participating middle schools to distribute recruitment flyers to students. Adolescents were asked to return the letter signed by their parents indicating whether the parents were interested in learning about the study. Parents who indicated interest were contacted by project staff. Families were eligible to participate if: (a) at least one parent was born in a Spanish-speaking country in the Americas, (b) the adolescent would advance to the eighth grade in the next school year, (c) neither the adolescent nor the primary parent had ever been hospitalized for psychiatric reasons, (d) the family was not planning to move out of the South Florida area during the 3 years of the study, (f) the adolescent was living with an adult primary caregiver who was willing to participate, and (g) a primary caregiver could attend weekly weekday evening meetings.

Of the 649 potential families, 70 declined to participate, 44 adolescents were not promoted to the eighth grade, 26 families were not living within the catchment areas of one of the three participating schools, 163 families reported scheduling conflicts, and 80 families were planning to move out of the area. The remaining 266 families met inclusion criteria and were randomized to one of the three intervention conditions. Each condition consisted of two modules. The conditions were: (a) Familias Unidas (family strengthening program) + Parent-Preadolescent Training for HIV Prevention (PATH), (b) English for Speakers of Other Languages (ESOL) + PATH, and (c) ESOL + HeartPower! for Hispanics (HEART-H), a cardiovascular health promotion program. All modules were delivered directly to parents in a participatory group format, and parents were encouraged to transmit the information to their adolescents in facilitator-supervised home visits. Intervention activities occurred between the baseline and 12-month assessment timepoints. More information on intervention outcomes is presented in Prado et al. (2007).

Participants were 128 boys and 138 girls ($M_{\text{age}} = 13.4$, $SD = 0.68$) and their primary caregivers (34 men, 232 women; $M_{\text{age}} = 40.9$, $SD = 6.2$). The vast majority (90%) of participating caregivers were the adolescents' biological mothers. Only 18.6% of the families reported household incomes greater than $30,000 per year. Nearly all (99%) of parents, and 60% of adolescents, were foreign-born. The remaining adolescents were born in the United States but raised by foreign-born parents. Immigrant adolescents and parents were born in Cuba (40%), Nicaragua (25%), Honduras (9%), Colombia (4%), and other Hispanic countries (22%). Of foreign-born adolescents ($n = 159$), 50% had been living in the United States for less than 3 years, whereas the other half had either been living in the United States between 3 and 10 years ($n = 54$; 34%) or more than 10 years ($n = 25$; 16%). Participants were assessed at baseline and at 6, 12, 24, and 36 months postbaseline. Of the original cohort of 266 families, 250 provided data at two or more timepoints, and 178 provided data at all five timepoints. Six families did not provide data at Time 2 but returned to the study at later timepoints. Individuals providing data at all timepoints differed from other participants only in terms of age. "Completers" were, on average, 13.4 years of age at baseline, compared to 13.6 years for other participants.

Adolescent measures were completed on laptop computers using the audio-computer-assisted-interviewing (A-CASI; Turner et al., 1998) system. The questions and response choices were read to the adolescent through a set of headphones connected to a laptop computer. The adolescent indicated her or his response using the keyboard or mouse. The A-CASI system has been shown to increase the veracity of reports of sensitive or illegal behaviors, especially in adolescents and young adults (Cooley et al., 2001). Parent assessments were completed in interview form with a trained Hispanic interviewer because, during the pilot phase, many parents expressed discomfort about completing their assessments on computer. Parents and adolescents were assured that their answers would be kept confidential to the extent permitted by law (i.e., except in cases of abuse or of imminent harm to oneself or to someone else). Each participant completed the study measures in the language of her or his choice. Nearly all (99%) parents completed their assessments in Spanish at all timepoints. At baseline, 57% of adolescents ($n = 151$) elected to complete their assessments in Spanish, and 39% ($n = 104$) elected to complete their assessments in English. We do not have language-of-assessment data for 11 adolescents (4.1% of the sample). Families were paid $25 at baseline, with payments increasing by $5 at each successive timepoint.

Measures

For each of the adolescent-reported measures used in our study (aside from the cigarette use, alcohol use, and sexual behavior measures, which were single-item scales), we evaluated measurement invariance across language of administration (see Knight, Roosa, & Umaña-Taylor, 2009, for a review of these procedures) using baseline data. Because almost all of the parents completed their assessments in Spanish, we were not able to ascertain measurement invariance in parent-reported measures. We sought to establish metric (equivalent factor loadings) and scalar (equivalent item intercepts) invariance for each measure across languages by comparing: first, a model with all factor loadings constrained equal across languages to a model with all factor loadings free to vary across languages, and second a model with all factor loadings constrained across languages to a model with all factor loadings and item intercepts constrained across languages. For measurement invariance to hold, the chi-square difference should be nonsignificant, and the differences in comparative fit index (CFI) and root mean square error of approximation (RMSEA) values should be .01 or less (Dimitrov, 2010; Widaman & Reise, 1997). Because the chi-square difference is often overpowered in measurement models (Chen, 2007), the assumption of invariance may be retained if the chi-square difference is significant but the differences in the other fit indices are less than .01. Although invariance testing can be extended to include residual variances, this is an overly restrictive assumption (Dimitrov, 2010; Vandenberg & Lance, 2000).

Acculturation. Parents and adolescents completed the Bicultural Involvement Questionnaire (Szapocznik et al., 1980). This 24-item measure asks participants about the extent to which they are comfortable engaging in a number of cultural practices (e.g., language use at home and at school or work, reading magazines, watching television, celebrating birthdays and holidays) in "American ways" and in "Hispanic ways." The response scale for each item ranges from 1 (*not at all comfortable*) to 5 (*very comfortable*). For the present analyses, we used acculturation data from Times 1–4.

Measurement invariance analyses on the BIQ Americanism and Hispanicism subscales indicated that neither the Americanism subscale, $\Delta\chi^2(18) = 38.40$, $p < .005$, $\Delta\text{CFI} = 0.015$, $\Delta\text{RMSEA} = .001$, nor the Hispanicism scale, $\Delta\chi^2(18) = 67.83$, $p < .001$, $\Delta\text{CFI} = 0.039$, $\Delta\text{RMSEA} = .013$, met criteria for linguistic measurement invariance. Examination of the factor loadings for the Hispanicism items indicated that the items referring to use of Spanish in various settings loaded well on the Hispanicism subscale factor for English-speaking adolescents, but loaded poorly for Spanish-speaking adolescents. No clear pattern emerged for the Americanism items.

Given that the language use items did not function equivalently across language of administration, we removed these items from the adolescent-reported Hispanicism and Americanism subscales in all further analyses. The adolescent-reported Hispanicism and Americanism subscales therefore consisted of seven items apiece. Cronbach's alpha coefficients for these shortened scales at baseline were .91 for Americanism and .88 for Hispanicism. The language items were retained for the parent-reported Americanism and Hispanicism subscales, because 99% of the parents completed their assessments in Spanish.

Family functioning. Family functioning was assessed in terms of parental involvement, positive parenting, and parent–adolescent communication. Parents and adolescents each completed separate measures of each of these constructs. *Parental involvement and positive parenting* were assessed using parent and adolescent versions of the Parenting Practices Scale (Gorman-Smith, Tolan, Zelli, & Huesmann, 1996). The adolescent-report parental involvement subscale consists of 12 items ($\alpha = .87$), the adolescent-report positive parenting subscale consists of 6 items ($\alpha = .91$), the parent-report involvement subscale consists of 10 items ($\alpha = .71$), and the parent-report positive parenting subscale consists of 6 items ($\alpha = .79$). Sample adolescent-report items include "How often did your parent ask you about your plans for the coming day?" (involvement) and "When you do something that your parent likes, does s/he give you a wink or a smile" (positive parenting). Aside from the two additional adolescent-reported involvement items, the parent-report items ask the same questions as the adolescent-report items, from the parent's perspective—such as "How often have you asked your child about her/his plans for the coming day?" The response scale for each item ranges from 1 (*almost never*) to 3 (*often*). *Parent–adolescent communication* was assessed using the Parent–Adolescent Communication Scale (Barnes & Olson, 1985). Both the parent and adolescent versions consist of 20 items (adolescent, $\alpha = .93$; parent, $\alpha = .81$). Sample items include "I can express my feelings to my parent/child without feeling restrained." The response scale ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). For the present analyses, family functioning data were used from Times 3 and 4.

Measurement invariance analyses on the family functioning measures indicated that parent involvement, positive parenting, and parent–adolescent communication met criteria for partial linguistic measurement invariance. Specifically, although we found full metric invariance, some of the item intercepts were not equivalent across languages for the parent involvement, positive parenting, and parent–adolescent communication measures. However, following Knight, Roosa, et al. (2009), we proceeded with these subscales given the assumption that the combination of full metric invariance and partial scalar invariance is sufficient to support pooling data across language of administration. Statistical details regarding these analyses are available from the first author on request.

Adolescent behavioral outcomes. Conduct problems, past-90-day cigarette and alcohol use, past-90-day sexual activity, and past-90-day unprotected sex at the Time 5 assessment were used as the outcome variables in this study. Time 4 scores on these same constructs were used to control for consistency in risky behaviors over time. Adolescent conduct problems were measured using shortened versions of the adolescent-reported aggressive behavior (12 items; $\alpha = .72$) and attention problems (9 items; $\alpha = .74$) subscales from the youth self-report (Achenbach & Rescorla, 2002). Sample items from these scales include "I am mean to others" (aggressive behavior) and "I have trouble paying attention" (attention problems). The response scale for each item ranges from 0 (*not true*) to 2 (*very often or often true*).

Measurement invariance analyses on the conduct problem measures indicated that although some measurement nonequivalence was present, $\Delta\chi^2(6) = 13.50$, $p = .036$, $\Delta\text{CFI} = 0.029$, $\Delta\text{RMSEA} = .020$, the fully constrained model provided acceptable model fit (CFI = 0.97, RMSEA = .095). Freeing one item intercept, as per modification indices, produced a nonsignificant chi-square difference between the constrained and unconstrained models, $\Delta\chi^2(5) = 4.55$, $p = .49$, and this revised constrained model fit the data extremely well (CFI = 1.00, RMSEA < .001). The behavior problem subscales therefore largely met criteria for linguistic measurement invariance.

Substance use. Substance use was assessed using items similar to those used in the Monitoring the Future Study (Johnston, O'Malley, Bachman, & Schulenberg, 2011). Adolescents were asked whether they had smoked in their lifetime and in the 90 days prior to the Time 4 and 5 assessments. Adolescents were also asked about how many occasions they had consumed alcohol in the 90 days prior to assessment. Response choices included 1 (*0 occasions*), 2 (*1–2 occasions*), 3 (*3–5 occasions*), 4 (*6–9 occasions*), 5 (*10–19 occasions*), 6 (*20–29 occasions*), and 7 (*40 + occasions*). Given low levels of response to Categories 3–7, we recoded the alcohol use response scale to 1 (*0 occasions*), 2 (*1–2 occasions*), or 3 (*3 + occasions*). We also asked about marijuana and hard drug use; however, because only nine adolescents reported any marijuana use, and only six adolescents reported any hard drug use in the 90 days prior to assessment, these substances were not included in the present analyses. Cigarette use was therefore measured as a dichotomous variable, and alcohol use was measured as an ordinal variable.

Sexual risk behaviors. Sexual risk behaviors were measured using items from Jemmott, Jemmott, and Fong's (1998) Sexual Behavior instrument. Adolescents were asked to indicate whether they had ever had sex (vaginal, anal, or oral), and for those who had ever had sex, whether they had had sex in the past 90 days preceding the Time 4 and Time 5 assessments. Adolescents who reported having had sex in the past 90 days were asked whether they had engaged in unprotected sex during that time period. Sexual outcomes were therefore measured as dichotomous variables.

Results

We present the study results in three general steps. First, we describe patterns and predictors of missingness in the study variables. Second, we create acculturation trajectory classes for adolescents and for parents. Finally, we estimate our proposed structural model (see Figure 1).

Step 1: Missing Data

We first examined patterns of missing data using Little's (1988) Missing Completely at Random

(MCAR) test. In cases where the MCAR assumption is violated, Enders (2010) recommends creating a missingness indicator for each variable (or set of variables) and comparing all other study variables between cases with and without missing data on the variable(s) in question. It should be noted that maximum likelihood estimation still outperforms more traditional ways of handling missing data (e.g., listwise deletion, mean replacement) even when data are not MCAR (Schlomer, Bauman, & Card, 2010). We conducted the MCAR test separately for the acculturation variables, family functioning variables, and outcome variables, because these variables were included in the study model at different timepoints. We included adolescent and parent reported variables together within each MCAR test. A normed chi-square (χ^2/df) of 1.5 or below indicates missingness completely at random (Enders, 2010).

For the acculturation variables at Times 1–4, the assumption of missingness completely at random was retained, $\chi^2(241) = 283.99$, $p < .05$, $\chi^2/df = 1.18$. For the family functioning variables at Times 3 and 4, the assumption of missingness completely at random was statistically rejected, $\chi^2(182) = 613.57$, $p < .001$, $\chi^2/df = 3.37$. For the adolescent outcome variables at Time 5, the assumption of missingness completely at random was also statistically rejected, $\chi^2(4) = 16.04$, $p < .005$, $\chi^2/df = 4.01$.

Because the family functioning and adolescent outcome variables did not satisfy the MCAR assumption, we examined differences in demographic variables and baseline scores on the study constructs between cases with and without valid family functioning data at Times 3 and 4. Of the 266 study families, 226 (85%) provided valid family functioning data at Time 3, and 213 (80%) provided valid family functioning data at Time 4. Multivariate analyses of variance (MANOVAs; one for Time 3 family functioning and one for Time 4 family functioning) indicated that none of the study variables at baseline were significantly related to missingness on family functioning at Times 3 or 4. Chi-square tests indicated that missingness on family functioning at Times 3 and 4 was not related to intervention condition or to adolescent gender or nativity (U.S.-born vs. foreign-born). We therefore proceeded with structural equation modeling. We used the mean and variance adjusted weighted least squares (WLSMV) estimator, which deletes missing data only for observed independent variables (Bengt Muthén, personal communication, October 3, 2005). Except for one adolescent who provided no acculturation data at any of the study timepoints, all adolescents and parents were assigned to an acculturation trajectory class (which was modeled as an observed predictor), and parent and adolescent reports of family functioning were modeled as latent variables, for which missing data are not deleted.

For the Time 5 adolescent outcome variables, 207 adolescents (78%) provided complete data. A MANOVA indicated no associations of the study variables at baseline with missingness at Time 5, Wilks's $\lambda = .95$, $F(10, 239) = 1.35$, $p = .20$, partial $\eta^2 = .05$. Chi-square analyses indicated no significant associations of intervention condition, adolescent gender, or adolescent nativity with missingness on adolescent outcomes at Time 5.

Step 2: Acculturation Trajectory Classes

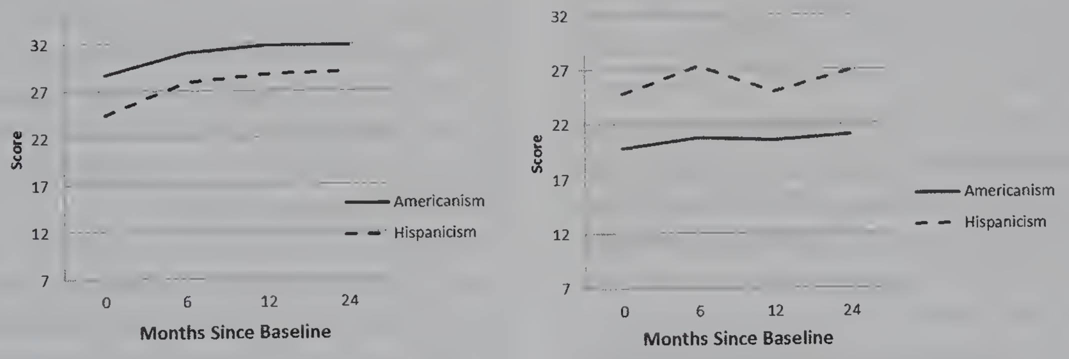
Our next step of analysis was to examine longitudinal trajectories of acculturation in terms of distinct classes of individuals grouped according to their baseline levels (at the end of seventh grade) and change trajectories (from seventh to ninth grades). We used a multistage decision process (Nylund, Asparouhov, & Muthén, 2007) to decide on the number of classes to retain by considering five criteria. First, the Vuong–Lo–Mendell–Rubin likelihood ratio test indicates the extent to which the −2 log likelihood value for a model with k classes is significantly smaller than the corresponding value for a model with $k - 1$ classes. Second, the Akaike information criterion and Bayesian information criterion provide an additional basis for comparing models, where lower values indicate better fit. Third, to ensure stability of the class solution, each class had to represent more than 5% of the sample. Fourth, classes had to be substantively different from one another (i.e., one class could not simply be a variant on another class). Fifth, posterior probabilities of correct classification and entropy values should be .75 or higher.

Growth curves for both Americanism and Hispanicism were included in the mixture model so that the solution extracted would be based on a bidimensional understanding of acculturation (cf. Nagin, 2005; Schwartz & Zamboanga, 2008). Separately for Americanism and for Hispanicism, the Time 1, 2, 3, and 4 scores were attached to latent intercept and slope terms, with all intercept loadings fixed to 1 and each slope loading fixed to the number of months elapsed since the baseline assessment (i.e., 0, 6, 12, and 24). We therefore extracted classes based on both Americanism and Hispanicism trajectories.

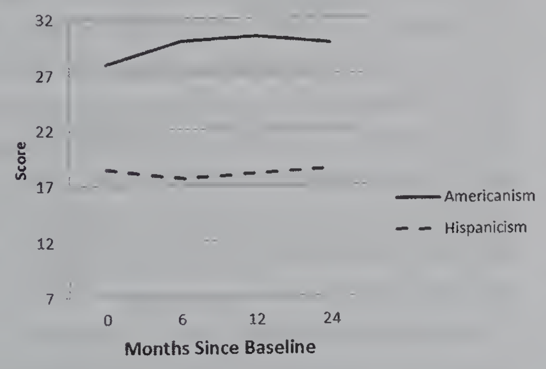
For adolescents, we extracted a three-class solution. The Vuong–Lo–Mendell–Rubin likelihood ratio test indicated a statistically significant improvement in fit when a third class was added, $LRT = 198.80$, $p < .02$, but not when a fourth class was added, $LRT = 67.50$, $p = .12$. The entropy value for the three-class solution was $E = .83$, and posterior classification accuracies for the three classes ranged from .89 to .95.

Based on the intercepts and slopes for adolescent Hispanicism and Americanism (see Figure 2), the three classes were named moderate bicultural ($n = 30$), high bicultural ($n = 171$), and assimilated ($n = 64$). Following Coatsworth et al. (2005), the scale midpoint, rather than the sample mean or median, was used to name the classes—as a way of limiting the sample specificity of the class labels. The moderate bicultural class was characterized by baseline Americanism and Hispanicism scores that were both close to the scale midpoint and did not change significantly over time. The assimilated class was characterized by Americanism scores near the top of the range of possible values, and by Hispanicism scores that begin near the scale midpoint. Neither Americanism nor Hispanicism scores changed significantly over time within the assimilated class. The high bicultural class was characterized by high and modestly increasing Americanism scores, and by Hispanicism scores that begin in the moderate-to-high range and increase markedly over time. Assimilated adolescents were most likely (53.1%) to be U.S.-born, whereas moderate bicultural adolescents were least likely (13.3%) to be U.S.-born, $\chi^2(2) = 13.49$, $p < .001$, Cramér's $V = .23$. Moreover, among foreign-born adolescents, those in the moderate bicultural class were most likely (73.1%) to have lived in the United States for 3 years or less, with participants in the assimilated (46.7%) and high bicultural (45.6%) classes marginally significantly less likely to be recent immigrants, $\chi^2(4) = 8.19$, $p = .085$, Cramér's $V = .16$.

We also extracted latent trajectory classes for parent Americanism and Hispanicism. The decision process pointed to a three-class solution, $LRT = 150.83$, $p < .001$, $E = .80$; posterior accuracy values ranged from .88 to .93. Class 1 (named highly Hispanic; $n = 153$) was characterized by extremely high Hispanicism scores and extremely low Americanism scores, Class 2 (named moderate bicultural; $n = 62$) was characterized by moderate-to-high Hispanicism scores and moderate Americanism scores, and Class 3 (named moderately hispanic; $n = 51$) was characterized by moderate Hispanicism scores and extremely low Americanism scores (see Figure 3). None of the slopes were significantly different from zero, indicating that parent acculturation scores changed little over time. The classes did not differ significantly in terms of number of years spent

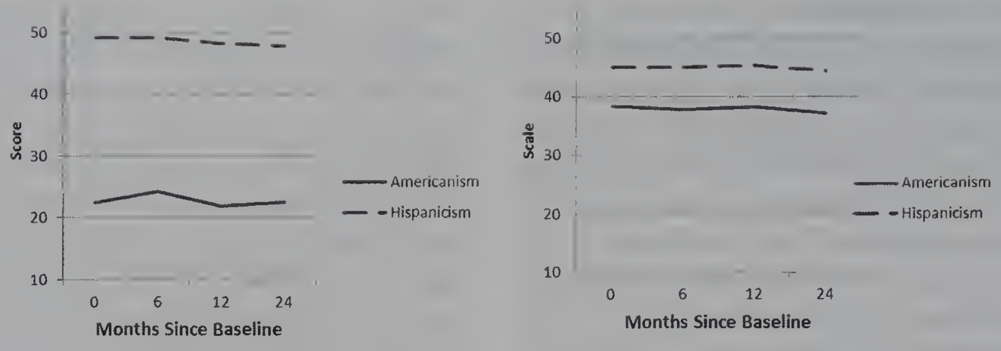


High Bicultural **Moderate Bicultural**



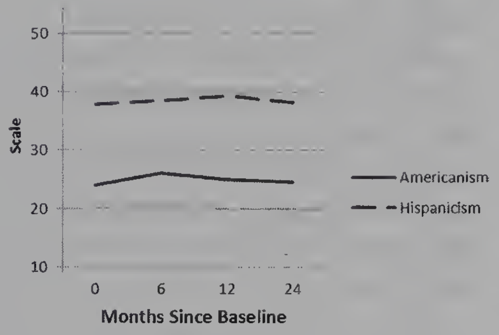
Assimilated

Figure 2. Adolescent acculturation class solution.



High Hispanic

Moderate Bicultural



Moderate Hispanic

Figure 3. Parent acculturation class solution.

in the United States. Not surprisingly, however, parents in the moderate bicultural class were most likely (51.6%) to have U.S.-born adolescents, $\chi^2(2) = 6.81$, $p < .04$, Cramér's $V = .16$. Parents in the highly Hispanic class were least likely (27.5%) to have U.S.-born adolescents.

Step 3: Acculturation as a Predictor of Adolescent Problem Outcomes

To test the primary study hypothesis—that trajectories of acculturation would predict adolescent outcomes both directly and indirectly through their effects on family functioning—we estimated a structural mediational model using the Mplus software package (release 6.12; Muthén & Muthén, 2011). To ensure that the mediational model was fully longitudinal, and to rigorously test the theoretical perspective advanced by Szapocznik and Kurtines (1993), acculturation trajectories were modeled in terms of the classes created from Time 1–4 scores, family functioning was modeled at Time 4, and adolescent behavioral outcomes were modeled at Time 5. To most rigorously evaluate mediation, we controlled for Time 3 family functioning so that the acculturation classes would be able to predict only the variability in Time 4 family functioning that was not associated with the stability of family functioning over time. Moreover, because adolescent and parent reports of family functioning are generally only modestly intercorrelated (Schwartz, Pantin, Prado, Sullivan, & Szapocznik, 2005), we created separate latent variables for parent and adolescent reports of family functioning at Times 3 and 4. Similarly, in predicting Time 5 adolescent outcomes, we controlled for the corresponding Time 4 adolescent outcome variables. Given that the acculturation trajectories were parameterized as ending at Time 4, and given that stability of family functioning from Times 3–4 was statistically controlled, the model provides a fairly rigorous longitudinal test of mediation (cf. Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001).

Frequencies and distributions of substance use and sexual behavior data. At the final study timepoint, 18 adolescents (9% of all adolescents providing data) indicated having smoked cigarettes during the 90 days prior to assessment, and 56 adolescents (27%) indicated having consumed alcohol at least once in the 3 months prior to assessment. Of the 207 adolescents providing sexual behavior data at the final study timepoint, 45 (21.7%) indicated having had anal or vaginal sex during the 90 days prior to assessment, and 30 (14.5%) indicated having had unprotected anal or vaginal sex during the 90 days prior to assessment.

Structural models. Four steps were followed in testing the structural model. First, we created latent variables for adolescent and parent family functioning. Second, we evaluated whether the structural model could be collapsed across intervention condition. Third, we examined the structural model. Finally, we tested the extent to which parent and adolescent reports of family functioning mediated the effects of parent and adolescent acculturation trajectory classes on adolescent outcomes.

In the structural model that we estimated (see Figure 1), adolescent reports of family functioning were parameterized as a latent variable with parent–adolescent communication, parental involvement, and positive parenting as indicators. Parent-reported family functioning used the same set of indicators. Fornell and Larcker (1981) proposed a method of calculating reliability for latent variables, where reliability is computed as the ratio of the variability explained by the latent variable to the total variability among the indicators. Reliability estimates were .83 and .82 for adolescent-reported family functioning at Times 3 and 4, respectively, and .59 and .73 for parent-reported family functioning at Times 3 and 4, respectively.

In estimating the structural model, we again used the WLSMV estimator, rather than the maximum likelihood estimator, because the WLSMV estimator does not require mathematical integration and provides standard fit indices for models with categorical outcomes. However, whereas maximum likelihood estimation provides logistic regression coefficients that can be converted to odds ratios, the WLSMV estimator provides probit regression coefficients (Azen & Walker, 2011). Probit regression assumes that the probability of the dependent-variable event occurring is normally distributed, and as such, a standardized probit regression coefficient is similar to a standardized linear regression coefficient (except that the outcome is the probability of event occurrence). As fit indices, we used the CFI and the RMSEA. Kline (2006) suggests that good model fit is represented by $\text{CFI} \geq 0.95$ and $\text{RMSEA} \leq .05$, and that adequate fit is represented by $\text{CFI} \geq 0.90$ and $\text{RMSEA} \leq .08$. The RMSEA also provides a 90% confidence interval and a “close fit probability” that the RMSEA value is below .05 (Hancock & Freeman, 2001).

Model invariance across conditions. To ensure that the sample could be collapsed across intervention conditions, we estimated the model within each condition and conducted equivalence analyses to ensure that the model fit similarly across conditions. The two control conditions were combined in the equivalence analysis to avoid extremely small cell sizes for some of the risk behavior outcomes. For example, the low frequency of cigarette smoking and risky sex at Time 5 created extremely small cell sizes when the sample was split by condition, and indeed the unconstrained model produced inadmissible parameter estimates (including correlations above 1.0). As a result, we conducted the invariance test without cigarette smoking and risky sex. Because the WLSMV estimator was used, the DIFFTEST option in Mplus was used to compare the constrained and unconstrained models to obtain a correct chi-square difference value.

To conduct the invariance test, we compared: (a) a fully unconstrained model with all factor loadings, structural paths, and intercepts free to vary across conditions with (b) a fully constrained model with all factor loadings, structural paths, and intercepts set equal across conditions. When testing invariance of structural paths and intercepts, Little (1997) recommends using only the chi-square difference index. The invariance test indicated that the model demonstrated metric (factor loadings and structural paths) and scalar (structural intercepts) invariance across conditions, $\Delta\chi^2(50) = 47.24$, $p = .58$. The constrained model provided an acceptable fit to the data, $\chi^2(432) = 490.83$, $\text{CFI} = 0.90$, $\text{RMSEA} = .036$, 90% CI [.014, .051].

Evaluating and interpreting the structural model. In the full model, because there were three adolescent acculturation trajectory classes and three parent acculturation trajectory classes, we used dummy coding to enter these classes as predictor variables. For adolescents, the high bicultural class was used as the reference group, and for parents, the high Hispanic class was used as the reference group—because these were the classes with the highest Hispanicism scores for each reporter. Dummy variables were created for each of the other two classes for each reporter, and were allowed to predict adolescent and parent reports of family functioning, as well as adolescent reports of behavior problems, cigarette and alcohol use, and sexual behavior, and risk taking. Where the effect of the acculturation classes on adolescent or parent reports of family functioning was statistically significant, we evaluated the possibility that family functioning might mediate the effects of acculturation trajectory on adolescent outcomes.

The structural model provided an acceptable fit to the data, $\chi^2(257) = 366.52$, $p < .001$, $\text{CFI} = 0.90$, $\text{RMSEA} = .045$, 90% CI [.034, .055], close fit probability = .79 (see Figure 4). Path coefficients for the dummy-coded acculturation class variables are reported as unstandardized coefficients, because dummy variables are on a nominal scale. A significant effect of acculturation trajectory class on family functioning emerged: The assimilated adolescent

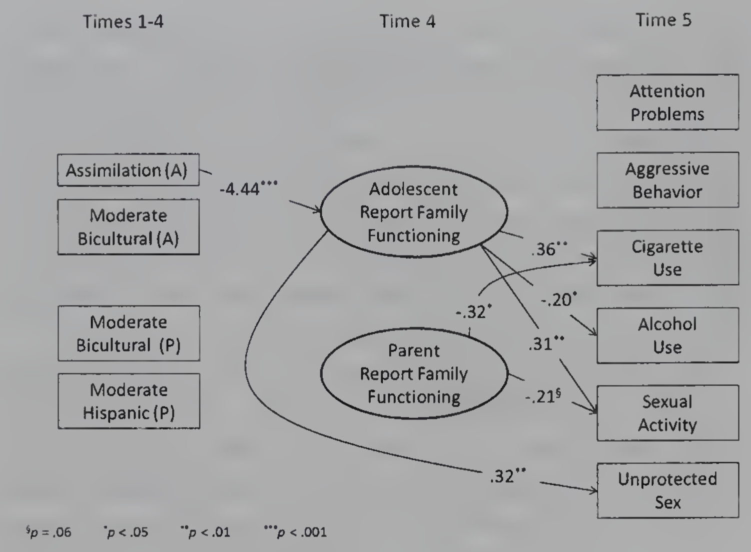


Figure 4. Final structural model.
Note. Only significant results are displayed. (A) = adolescent; (P) = parent.

class was associated with less favorable adolescent-reported family functioning compared to the high bicultural class, $B = -4.44$, $p < .02$. A marginally significant direct effect of trajectory class on adolescent outcomes emerged. Compared to those in the high bicultural class, adolescents in the assimilated class reported marginally higher levels of aggressive behavior, $B = 1.11$, $p = .068$.

A number of associations emerged between family functioning at Time 4 and adolescent outcomes at Time 5. Parent-reported family functioning was significantly negatively predictive of adolescent cigarette smoking, $\beta = -.32$, $p < .02$, and marginally significantly predictive of sexual activity, $\beta = -.21$, $p = .062$. Adolescent-reported family functioning was negatively predictive of extent of alcohol use, $\beta = -.20$, $p < .04$, but was positively predictive of cigarette smoking, $\beta = .36$, $p < .02$; sexual activity, $\beta = .31$, $p < .01$; and unprotected sex, $\beta = .32$, $p < .01$.

Tests of mediation. Because adolescent and parent reports of family functioning were associated both with acculturation classes and with adolescent outcomes, we tested the extent to which family functioning would mediate the effects of acculturation trajectory classes on adolescent outcomes. Such mediated associations would suggest that adolescents' or parents' classes (or patterns) of Americanism and Hispanicism over time would affect adolescent behavioral outcomes through adolescent or parent perceptions of family functioning. We tested for mediation using MacKinnon's (2008) Asymmetric Distribution of Products Test and using the Model Indirect command in Mplus (Muthén & Muthén, 2011). This test computes a 95% confidence interval around the product of the two unstandardized path coefficients that comprise the mediating pathway. If this confidence interval does not include zero, then partial mediation is assumed. A standardized indirect path coefficient can be computed by multiplying the two standardized path coefficients that comprise the mediating pathway.

Three significant mediated pathways emerged: (a) a negative effect of the assimilated adolescent class on cigarette smoking through adolescent-reported family functioning, unstandardized $B = -.193$, 95% CI [−.382, −.003], $p < .05$, standardized $\beta = -.075$; (b) a negative effect of the assimilated adolescent class on sexual activity through adolescent-reported family functioning, unstandardized $B = -.168$, 95% CI [−.324, −.012], $p < .04$, standardized $\beta = -.065$; and (c) a negative effect of the assimilated adolescent class on risky sexual activity through adolescent-reported family functioning, unstandardized $B = -.164$, 95% CI [−.328, −.001], $p = .05$, standardized $\beta = -.067$.

Follow-up analysis: Examining the counterintuitive effects of adolescent-reported family functioning. The finding that adolescent-reported family functioning positively predicted adolescent cigarette smoking, sexual activity, and unprotected sex appears counterintuitive and warrants further examination. Because both parent and adolescent reports of family functioning were included as predictors of adolescent outcomes, the effects of adolescent family functioning reports are calculated controlling for the effects of parent family functioning reports (and vice versa; Keith, 2006). As a result, the counterintuitive effects of adolescent-reported family functioning on adolescent outcomes may have been a result of discrepancies between parent and adolescent reports of family functioning.

To explore this possibility, we created a latent parent–adolescent difference score (adolescent report minus parent report; King et al., 2006) at Times 3 and 4, and we used this difference score in place of adolescent and parent reports of family functioning. The model including the difference score fit the data adequately, $\chi^2(255) = 353.36$, $p < .001$, CFI = 0.91, RMSEA = .043, 90% CI [.031, .053], close fit probability = .87. Neither parent nor adolescent acculturation trajectories significantly predicted the latent family functioning difference score at Time 4 (controlling for the latent difference score at Time 3). However, the latent difference score at Time 4 significantly predicted adolescent outcomes at Time 5: cigarette smoking, $\beta = .34$, $p < .01$; alcohol use, $\beta = -.19$, $p < .04$; sexual activity, $\beta = .27$, $p < .005$; and unprotected sex, $\beta = .27$, $p < .01$.

Discussion

This study examined the developmental trajectories of acculturation in Hispanic adolescents and parents, as well as the effects of these trajectories on family functioning and on adolescent behavioral outcomes. Several major findings emerged. First, of the three adolescent trajectory classes that we extracted, only the high bicultural class was characterized by changes (increases) in Americanism and Hispanicism over time. However, 64% of adolescents were placed into the high bicultural class, suggesting that the majority of adolescents in the sample were high and increasing in both Americanism and Hispanicism. This pattern suggests that the bicultural environment of Miami is likely to enrich both American and Hispanic cultural streams within Hispanic adolescents who are already bicultural—but not necessarily in those who adopt other approaches to acculturation. Such a finding is consistent with the supposition that a bicultural environment can encourage and increase biculturalism in young people (cf. Flannery, Reise, & Yu, 2001; Schwartz & Zamboanga, 2008). Miami appears to represent an example of a bicultural environment that facilitates the continuing development of cultural behaviors for the majority of Hispanic adolescents—and some other U.S. cities with large Hispanic populations (such as Los Angeles, New York, Chicago, and Houston) may provide similarly bicultural environments.

Second, as an index of validation for the trajectory classes, these classes mapped as expected onto adolescents' nativity and time in the United States. The moderate bicultural class, which appeared to represent a combination of Berry's (1980, 1997) separated and bicultural approaches, consisted primarily of recently immigrated adolescents—whereas U.S.-born adolescents were overrepresented within the assimilated class. Although parents' length of time (less than 3 years, 3–10 years, or more than 10 years) in the United States was not related to their acculturation trajectory classification, those parents whose adolescents were born in the United States were disproportionately likely to be classified into the moderate bicultural trajectory class, which was characterized by higher levels of Americanism compared to the other two classes. Hispanicism was moderate to high in all three parent trajectory classes, suggesting that the Hispanic enclave context of Miami was successful in maintaining Hispanic cultural practices among adults. Moreover, the high Hispanicism in all three parent trajectory classes may explain why these classes were not differentially related to parent reports of family functioning. Because many Hispanic cultures are characterized by a strong orientation toward family (Sabogal, Marin, Otero-Sabogal, Marin, & Perez-Stable, 1987), maintaining Hispanic cultural orientations may help to preserve parent–adolescent relationships—at least from the parent's perspective.

Third, the high bicultural adolescent acculturation trajectory—high and increasing in both Hispanicism and Americanism over time—was associated with the most favorable adolescent-reported family functioning. Because the highest degrees of biculturalism involve a deepening of one's solidarity both with the culture of settlement and with one's cultural heritage (Benet-Martínez & Haritatos, 2005), high bicultural adolescents may be in an advantaged position in terms of relating to traditionally oriented parents. Bicultural adolescents may also serve as a conduit between family members and some services within the culture of settlement (e.g., medical providers, bill collectors; Morales & Hanson, 2005)—although this "language brokering" role may have negative as well as positive consequences for adolescents (Trickett & Jones, 2007).

With that said, however, an assimilated—rather than highly bicultural—adolescent acculturation trajectory was indirectly protective against adolescent cigarette smoking and sexual activity, with adolescent-reported family functioning serving as a mediator. One would expect a bicultural approach to be predictive of the lowest levels of risk taking—and these counterintuitive mediated effects appeared to be driven by the positive link between adolescents' perceptions of family functioning and their likelihood of smoking cigarettes and engaging in sexual activity and in unprotected sex. Furthermore, exploring this anomalous finding indicated that the significant positive effects of adolescent-reported family functioning on risky outcomes may have been due to the divergence between adolescent and parent reports of family functioning. Specifically, in families where adolescents perceived family functioning to be more positive than parents did, adolescents appeared to be at greatest risk for cigarette smoking, sexual activity, and unprotected sex. However, the acculturation trajectory classes did not predict this discrepancy in reports of family functioning. Further research is needed to examine the effects of adolescent acculturation on family functioning over time, as well as the extent to which the *interaction* between parent and adolescent acculturation trajectories may affect family and adolescent outcomes.

Nonetheless, the negative effect of the assimilated adolescent acculturation trajectory on adolescent reports of family functioning is consistent with prior research suggesting links between loss of one's cultural heritage and family conflict and disengagement (Gil et al., 2000; Gonzales et al., 2006; McQueen et al., 2003). Indeed, a primary characteristic of assimilation is rejecting one's cultural heritage (Berry, 1980, 1997). Contrary to unidimensional approaches to acculturation, adopting American practices and customs is not problematic as long as Hispanic practices and customs are retained. Indeed, the more positive adolescent-reported family relationships associated with the two bicultural classes suggest that biculturalism represents an especially favorable approach to acculturation.

Fourth, parent-reported family functioning appeared to be protective against adolescent cigarette smoking and sexual activity. These effects may be a function of parental involvement and communication. Research indicates that Hispanic parents who are able to communicate with their adolescents in general may be more likely to discuss sexuality, smoking, and other risk behaviors with them (Harakeh, Scholte, Vermulst, de Vries, & Engels, 2010; McKee & Karasz, 2006). In turn, parent–adolescent communication about cigarette smoking is predictive of lower likelihood of adolescent smoking (de Leeuw, Scholte, Vermulst, & Engels, 2010), and parent–adolescent communication about sexuality is predictive of delayed sexual onset, and decreased likelihood of unprotected sex, among adolescents (Aspy et al., 2007).

Finally, all of the adolescent substance use and sexual outcomes were associated (directly, indirectly, or both) with parent or adolescent reports of family functioning. This suggests that the present findings were likely not due to shared-method biases involved in analyzing variables all reported by the same person (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Favorable adolescent perceptions of family functioning were protective against alcohol use, and favorable parent reports of family functioning were protective against cigarette smoking and recent sexual activity. The finding that parent, but not adolescent, perceptions of family functioning were protective against cigarette smoking and sexual activity is noteworthy. A likely explanation, as noted above, is that warm and close parent–adolescent relationships facilitate adolescent disclosure, parental guidance, and parental attempts to transmit their values regarding risky behaviors (Kerr & Stattin, 2000). Moreover, the extent to which parents and adolescents agree on the affective valence of their relationships with one another may also help to protect adolescents against cigarette smoking and precocious sexual activity.

Strengths and Limitations of This Study

Despite their contributions to the literature, the present results should be interpreted in light of several important limitations. Perhaps the most important limitation is that the present data were drawn from an intervention data set. Although there is some precedent for conducting longitudinal analyses on intervention data sets (e.g., Choi et al., 2005; Matsunaga et al., 2010), and although the model paths were not moderated by condition, sample composition is likely somewhat different between studies designed to test interventions and those designed purely to follow cohorts over time. Specifically, individuals and families who are most interested in receiving help (Perrino, Coatsworth, Briones, Pantin, & Szapocznik, 2001) are most likely to be engaged in an intervention study. However, it is noteworthy that 94% of families provided data at two or more timepoints, and 78% of families provided data at Time 5. This high retention rate suggests that we were able to track developmental change for the overwhelming majority of adolescents in our sample.

Second, although theory (Szapocznik & Kurtines, 1993) suggests that acculturation affects family functioning and adolescent outcomes (both directly and indirectly through family functioning), it is possible that family functioning and adolescent outcomes may also influence the trajectory of acculturation. Acknowledging this possibility is especially important given that family functioning (using either adolescent or parent reports) did not significantly predict conduct problems once earlier levels of conduct problems were controlled. It is important for future research to empirically test the directionality of the links between acculturative and family processes.

A third limitation concerns the use of a Miami Hispanic sample. Miami is a unique cultural context in at least two respects. First, it is the only major U.S. city where Hispanics hold the majority of political and economic power, and second, it is the only U.S. city where neither of the two largest Hispanic nationalities in the United States (Mexican Americans and Puerto Ricans) are well represented within the local Hispanic population. As a result, it is important to replicate the present findings across multiple geographic contexts to ascertain generalizability. We (Schwartz et al., 2012) are beginning to conduct research in this direction.

Fourth, because the language use items did not function equivalently across language of administration, we deleted those items and estimated the latent class growth analyses and structural models using only the items assessing other cultural practices (e.g., media use, choice of friends). There is evidence that language use is empirically distinguishable from other types of cultural practices (Kang, 2006). This suggests that the omission of language use items may have produced different results than might have emerged had language use items been included in analysis. However, our results also underscore the importance of examining linguistic measurement equivalence when measuring acculturation—and our findings suggest that language use items may operate differently in Spanish-dominant Hispanics than in Hispanics who are comfortable in English. Future research should further examine the different meanings of language use—and the role of language use as a dimension of acculturation—in bilingual versus monolingual Hispanic individuals.

Fifth, the use of self-reports of substance use and sexual behavior represents a potential limitation. Adolescents may overreport or underreport these behaviors for any number of reasons (cf. Shillington & Clapp, 2000). However, it should be noted that high levels convergence have been found between self-reports and biological measures of substance use among Hispanic adolescents (Dillon, Turner, Robbins, & Szapocznik, 2005), and that the method for data collection, the A-CASI, has been found to increase adolescents' veracity of responding (Cooley et al., 2001).

Finally, the dichotomous (yes or no) response scale used to assess cigarette use, sexual behavior, and sexual risk taking did not permit us to examine the extent or frequency of engagement in these activities. It is important for future research to obtain precise estimates of how often, and to what extent, adolescents engaged in specific risky behaviors.

In conclusion, and despite these limitations, this study has extended the literature on acculturation trajectories in Hispanic families, and on the associations of these trajectories with family functioning and risk behaviors. The finding that both parent and adolescent acculturation trajectories—and both adolescent and parent reports of family functioning—influenced adolescent risk behavior is consistent with a family-systemic perspective on adolescent outcomes. It is hoped that the present results inspire additional longitudinal research examining trajectories of acculturation and of their effects on family functioning and risk outcomes in Hispanic adolescents.

References

Achenbach, T. M., & Rescorla, L. A. (2002). *Manual for the ASEBA adult forms and profiles*. Burlington: Research Center for Children, Youth, and Families, University of Vermont.

Allen, M., Elliott, M., Fuligni, A., Morales, L., Hambarsoomian, K., & Schuster, M. (2008). The relationship between Spanish language use and substance use behaviors among Latino youth: A social network approach. *Journal of Adolescent Health, 43*, 372–379. doi:10.1016/j.jadohealth.2008.02.016

Aspy, C. B., Vesely, S. K., Oman, R. F., Rodine, S., Marshall, L., & McLeroy, K. (2007). Parental communication and youth sexual behavior. *Journal of Adolescence, 30*, 449–466. doi:10.1016/j.adolescence.2006.04.007

Azen, R., & Walker, C. M. (2011). *Categorical data analysis for the behavioral and social sciences*. New York: Routledge.

Barnes, H. L., & Olson, D. H. (1985). Parent-adolescent communication and the circumplex model. *Child Development, 56*, 438–447.

Benet-Martínez, V., & Haritatos, J. (2005). Bicultural identity integration (BII): Components and psychosocial antecedents. *Journal of Personality, 73*, 1015–1050. doi:10.1111/j.1467-6494.2005.00337.x

Berkel, C., Knight, G. P., Zeiders, K. H., Tein, J.-Y., Roosa, M. W., Gonzales, N. A., et al. (2010). Discrimination and adjustment for Mexican American adolescents: A prospective examination of the benefits of culturally related values. *Journal of Research on Adolescence, 20*, 893–915. doi:10.1111/j.1532-7795.2010.00668.x

Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. M. Padilla (Ed.), *Acculturation: Theory, models, and some new findings* (pp. 9–25). Boulder, CO: Westview.

Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review, 46*, 5–34.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 15*, 464–504. doi:10.1080/10705510701301834

Choi, Y., Harachi, T. W., Gillmore, M., & Catalano, R. F. (2005). Applicability of the social development model to urban ethnic minority youth: Examining the relationship between external constraints, family socialization, and problem behaviors. *Journal of Research on Adolescence, 15*, 505–534. doi:10.1111/j.1532-7795. 2005. 00109.x

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of*

Consulting and Clinical Psychology, 70, 6–20. doi:10.1037/0022-006X.70.1.6

Coatsworth, J. D., Maldonado-Molina, M. M., Pantin, H., & Szapocznik, J. (2005). A person-centered and ecological investigation of acculturation strategies in Hispanic immigrant youth. *Journal of Community Psychology, 33*, 157–174. doi:10.1002/jcop.20046

Cooley, P. C., Rogers, S. M., Turner, C. F., Al-Tayyib, A. A., Willis, G., & Ganapathi, L. (2001). Using touch screen audio-CASI to obtain data on sensitive topics. *Computers in Human Behavior, 17*, 285–293.

Crockett, L. J., & Zamboanga, B. L. (2009). Latino substance use: Cultural, social, and psychological considerations. In F. Villaruel, G. Carlo, M. Azmitia, J. Grau, N. Cabrera, & J. Chahin (Eds.), *Handbook of Latino psychology* (pp. 379–398). Newbury Park, CA: Sage.

de Leeuw, R., Scholte, R., Vermulst, A., & Engels, R. (2010). The relation between smoking-specific parenting and smoking trajectories of adolescents: How are changes in parenting related to changes in smoking? *Psychology and Health, 25*, 999–1021. doi:10.1080/08870440903477204

Dillon, F. R., De La Rosa, M., Schwartz, S. J., Rojas, P., Duan, R., & Malow, R. M. (2010). U.S. Latina age of sexual debut: Long-term associations and implications for HIV and drug abuse prevention. *AIDS Care, 22*, 431–440. doi:10.1080/09540120903202871

Dillon, F. R., Turner, C. W., Robbins, M. S., & Szapocznik, J. (2005). Concordance among biological, interview, and self-report measures of drug use among African American and Hispanic adolescents referred for drug abuse treatment. *Psychology of Addictive Behaviors, 19*, 404–413. doi:10.1037/0893-164X.19.4.404

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121–149. doi:10.1177/0748175610373459

Dinh, K. H., Roosa, M. W., Tein, J.-Y., & Lopez, V. A. (2002). The relationship between acculturation and problem behavior proneness in a Hispanic youth sample: A longitudinal mediation model. *Journal of Abnormal Child Psychology, 30*, 295–309.

Dumka, L. E., Gonzales, N. A., Bonds, D. D., & Millsap, R. E. (2009). Academic success of Mexican origin adolescent boys and girls: The role of mothers' and fathers' parenting and cultural orientation. *Sex Roles, 60*, 588–599. doi:10.1007/s11199-008-9518-z

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.

Flannery, W. P., Reise, S. P., & Yu, J. (2001). A comparison of acculturation models. *Personality and Social Psychology Bulletin, 27*, 1035–1045.

Fornell, C. R., & Larcker, D. F. (1981). Two structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39–50.

Fuligni, A. J. (2001). A comparative longitudinal approach to acculturation among children from immigrant families. *Harvard Educational Review, 71*, 566–578.

Gil, A. G., Wagner, E. F., & Vega, W. A. (2000). Acculturation, familism, and alcohol use among Latino adolescent males: Longitudinal relations. *Journal of Community Psychology, 28*, 443–458.

Gonzales, N. A., Deardorff, J. M., Formoso, D., Barr, A., & Barrera, M., Jr. (2006). Family mediators of the relation between acculturation and adolescent mental health. *Family Relations, 55*, 318–330. doi:10.1111/j.1741-3729.2006.00405.x

Gorman-Smith, D., Tolan, P. H., Zelli, A., & Huesmann, L. R. (1996). The relation of family functioning to violence among inner-city minority youths. *Journal of Family Psychology, 10*, 115–129.

Guilamo-Ramos, V., Jaccard, J. J., Peña, J., & Goldberg, V. (2005). Acculturation-related variables, sexual initiation, and subsequent sexual behavior among Puerto Rican, Mexican, and Cuban youth. *Health Psychology, 24*, 88–95. doi:10.1037/0278-6133.24.1.88

Hancock, G. R., & Freeman, M. J. (2001). Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educational and Psychological Measurement, 61*, 741–758.

Harakeh, Z., Scholte, R. H. J., Vermulst, A. A., de Vries, H., & Engels, R. C. M. E. (2010). The relations between parents' smoking, general parenting, parental smoking communication, and adolescents' smoking. *Journal of Research on Adolescence, 20*, 140–165. doi:10.1111/j.1532-7795.2009.00626.x

Jemmott, J. B., III, Jemmott, L. S., & Fong, G. T. (1998). Abstinence and safer sex HIV risk-reduction interventions for African American adolescents: A randomized controlled trial. *Journal of the American Medical Association, 279*, 1529–1536.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2011). *Monitoring the Future national survey results on drug use, 1975–2010: Vol. 1. Secondary school students*. Ann Arbor, University of Michigan, Institute for Social Research.

Kang, S. M. (2006). Measurement of acculturation, scale formats, and language competence: Their implications for adjustment. *Journal of Cross-Cultural Psychology, 37*, 669–693. doi:10.1177/0022022106292077

Keith, T. Z. (2006). *Multiple regression and beyond*. Boston: Allyn & Bacon.

Kepka, D., Coronado, G., Rodriguez, H., & Thompson, B. (2010). Acculturation and HPV infection among Latinas in the United States. *Preventive Medicine, 51*, 182–184. doi:10.1016/j.ypmed.2010.06.002

Kerr, M., & Stattin, H. (2000). What parents know, how they know it, and several forms of adolescent adjustment: Further support for a reinterpretation of monitoring. *Developmental Psychology, 36*, 366-380.

King, L. A., King, D. W., McArdle, J. J., Saxe, G. N., Doron-LaMarca, S., & Orazem, R. J. (2006). Latent difference score approach to longitudinal trauma research. *Journal of Traumatic Stress, 19*, 771–785. doi:10.1002/jts.20188

Kline, R. B. (2006). *Principles and practices of structural equation modeling* (2nd ed.). New York: Guilford.

Knight, G. P., Roosa, M. W., & Umaña-Taylor, A. J. (2009). Introduction to studying ethnic minority and economically disadvantaged populations. In G. P. Knight, M. W. Roosa, & A. J. Umaña-Taylor (Eds.), *Studying ethnic minority and economically disadvantaged populations: Methodological challenges and best practices* (pp. 3–28). Washington, DC: American Psychological Association.

Knight, G. P., Vargas-Chanes, D., Losoya, S. H., Cota-Robles, S., Chassin, L., & Lee, J. (2009). Acculturation and enculturation trajectories among Mexican-American adolescent offenders. *Journal of Research on Adolescence, 19*, 625–653. doi:10.1111/j.1532-7795.2009.00614.x

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158*, 848–856.

Lee, J., & Hahm, H. C. (2010). Acculturation and sexual risk behaviors among Latina adolescents transitioning to young adulthood. *Journal of Youth and Adolescence, 39*, 414–427. doi:10.1007/s10964-009-9495-8

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198–1202.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.

Matsunaga, M., Hecht, M., Elek, E., & Ndiaye, K. (2010). Ethnic identity development and acculturation: A longitudinal analysis of Mexican-heritage youth in the Southwest United States. *Journal of Cross-Cultural Psychology, 41*, 410–427. doi:10.1177/0022022109359689

McKee, M., & Karasz, A. (2006). "You have to give her that confidence": Conversations about sex in Hispanic mother-daughter dyads. *Journal of Adolescent Research, 21*, 158–184. doi:10.1177/0743558405285493

McQueen, A., Getz, J. G., & Bray, J. H. (2003). Acculturation, substance use, and deviant behavior: Examining separation and family conflict as mediators. *Child Development, 74*, 1737–1750. doi:10.1046/j.1467-8624.2003.00635.x

Mogro-Wilson, C. (2008). The influence of parental warmth and control on Latino adolescent alcohol use. *Hispanic Journal of Behavioral Sciences, 30*, 89–105. doi:10.1177/0739986307310881

Morales, A., & Hanson, W. E. (2005). Language brokering: An integrative review of the literature. *Hispanic Journal of Behavioral Sciences, 27*, 471–503. doi:10.1177/0739986305281333

Muthén, B. O., & Muthén, L. K. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. *Journal of Studies on Alcohol, 61*, 290–300.

Muthén, L. K., & Muthén, B. O. (2011). *Mplus user's guide version 6*. Los Angeles: Muthén and Muthén.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569. doi:10.1080/10705510701575396

Pahl, K., & Way, N. (2006). Longitudinal trajectories of ethnic identity among urban Black and Latino adolescents. *Child Development, 77*, 1403–1415. doi:10.1111/j.1467-8624.2006.00943.x

Pantin, H., Coatsworth, J. D., Feaster, D. J., Newman, F. L., Briones, E., Prado, G., et al. (2003). Familias Unidas: Efficacy of an intervention to promote parental investment in Hispanic immigrant families. *Prevention Science, 4*, 189–201. doi:10.1023/A:1024601906942

Pantin, H., Prado, G., Lopez, B., Huang, S., Tapia, M., Schwartz, S. J., et al. (2009). A randomized trial of Familias Unidas for Hispanic behavior-problem adolescents. *Psychosomatic Medicine, 71*, 987–995. doi:10.1097/PSY.0b013e3181bb2913

Park, S., & Romer, D. (2010). Problem behavior syndrome and its influence on the transition from experimental to daily smoking among adolescents in the U.S. *Journal of Addictions Nursing, 21*, 114–122. doi:10.3109/10884602.2010.481506

Perrino, T., Coatsworth, J. D., Briones, E., Pantin, H., & Szapocznik, J. (2001). Initial engagement in parent-centered preventive interventions: A family systems perspective. *Journal of Primary Prevention*, 22, 21–44.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903. doi:10.1037/0021-9010.88.5.879

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. Berkeley: University of California Press.

Portes, A., & Rumbaut, R. G. (2006). *Immigrant America: A portrait*, 3rd ed. Berkeley: University of California Press.

Prado, G., Pantin, H., Briones, E., Schwartz, S. J., Feaster, D. J., Huang, S., et al. (2007). A randomized controlled trial of a family-centered intervention in preventing substance use and HIV risk behaviors in Hispanic adolescents. *Journal of Consulting and Clinical Psychology, 75*, 914–926. doi:10.1037/0022-006X.75.6.914

Ramirez, J., Crano, W., Quist, R., Burgoon, M., Alvaro, E., & Grandpre, J. (2004). Acculturation, familism, parental monitoring, and knowledge as predictors of marijuana and inhalant use in adolescents. *Psychology of Addictive Behaviors, 18*, 3–11. doi:10.1037/0893-164X.18.1.3

Rudmin, F. W. (2003). Critical history of the acculturation psychology of assimilation, separation, integration, and marginalization. *Review of General Psychology, 7*, 3–37. doi:10.1037/1089-2680.7.1.3

Ryder, A., Alden, L., & Paulhus, D. (2000). Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology*, *79*, 49–65.

Sabogal, F., Marin, G., Otero-Sabogal, R., Marin, B. V., & Perez-Stable, E. (1987). Hispanic familism and acculturation: What changes and what doesn't? *Hispanic Journal of Behavioral Sciences*, *9*, 397–412.

Santisteban, D. A., & Mitrani, V. B. (2003). The influence of acculturation processes on the family. In K. M. Chun, P. B. Organista, & G. Marín (Eds.), *Acculturation: Advances in theory, measurement, and applied research*. Washington, DC: American Psychological Association.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*, 1–10.

Schwartz, S. J., Pantin, H., Prado, G., Sullivan, S., & Szapocznik, J. (2005). Family functioning, identity, and problem behavior in Hispanic immigrant early adolescents. *Journal of Early Adolescence*, *25*, 392–420. doi:10.1177/0272431605279843

Schwartz, S. J., Pantin, H., Sullivan, S., Prado, G., & Szapocznik, J. (2006). Nativity and years in the receiving culture as markers of acculturation in ethnic enclaves. *Journal of Cross-Cultural Psychology*, *37*, 345–353. doi:10.1177/0022022106286928

Schwartz, S. J., Unger, J. B., Des Rosiers, S. E., Huang, S., Baezconde-Garbanati, L., Lorenzo-Blanco, E., et al. (2012). Substance use and sexual behavior among recent Hispanic immigrant adolescents: Effects of parent-adolescent differential acculturation and communication. *Drug and Alcohol Dependence*, *125S*, S26–S34. doi:10.1016/j.drugalcdep.2012.05.020

Schwartz, S. J., Unger, J. B., Zamboanga, B. L., & Szapocznik, J. (2010). Rethinking the concept of acculturation: Implications for theory and research. *American Psychologist*, *65*, 237–251.

Schwartz, S. J., & Zamboanga, B. L. (2008). Testing Berry's model of acculturation: A confirmatory latent class approach. *Cultural Diversity and Ethnic Minority Psychology*, *14*, 275–285. doi:10.1037/a0012818

Shillington, A., & Clapp, J. (2000). Self-report stability of adolescent substance use: Are there differences for gender, ethnicity and age? *Drug and Alcohol Dependence*, *60*, 19–27.

Smokowski, P. R., & Bacallao, M. (2011). *Becoming bicultural: Risk, resilience, and Latino youth*. New York: New York University Press.

Smokowski, P. R., Rose, R., & Bacallao, M. (2008). Acculturation and Latino family processes: How cultural involvement, biculturalism, and acculturation gaps influence family dynamics. *Family Relations*, *57*, 295–308. doi:10.1111/j.1741-3729.2008.00501.x

Steiner, N. (2009). *International migration and citizenship today*. New York: Routledge.

Stepick, A., & Stepick, C. D. (2002). Power and identity: Miami Cubans. In M. M. Suárez-Orozco, & M. Páez (Eds.), *Latinos: Remaking America* (pp. 75–92). Cambridge, MA: Harvard University Press.

Szapocznik, J., & Kurtines, W. M. (1993). Family psychology and cultural diversity: Opportunities for theory, research, and application. *American Psychologist*, *48*, 400–407.

Szapocznik, J., Kurtines, W. M., & Fernandez, T. (1980). Biculturalism and adjustment among Hispanic youths. *International Journal of Intercultural Relations*, *4*, 354–375.

Szapocznik, J., Santisteban, D., Rio, A., Perez-Vidal, A., Santisteban, D. A., & Kurtines, W. M. (1989). Family effectiveness training: An intervention to prevent drug abuse and problem behavior in Hispanic adolescents. *Hispanic Journal of Behavioral Sciences*, *11*, 3–27.

Trickett, E. J., & Jones, C. J. (2007). Adolescent culture brokering and family functioning: A study of families from Vietnam. *Cultural Diversity and Ethnic Minority Psychology*, *13*, 143–150. doi:10.1037/1099-9809.13.2.143

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. B., Pleck, J. H., & Sonsenstein, L. H. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, *280*, 867–873.

Umaña-Taylor, A. J., Bhanot, R., & Shin, N. (2006). Ethnic identity formation during adolescence: The critical role of families. *Journal of Family Issues*, *27*, 390–414. doi:10.1177/0192513X05282960

Unger, J. B., Gallaher, P., Shakib, S., Ritt-Olson, A., Palmer, P., & Johnson, C. A. (2002). The AHIMSA acculturation scale: A new measure of acculturation for adolescents in a multicultural society. *Journal of Early Adolescence*, *22*, 225–251.

Unger, J. B., Ritt-Olson, A., Soto, D. W., & Baezconde-Garbanati, L. (2009). Parent-child acculturation discrepancies as a risk factor for substance use among Hispanic adolescents in Southern California. *Journal of Immigrant and Minority Health*, *11*, 149–157. doi:10.1007/s10903-007-9083-5

Unger, J. B., Ritt-Olson, A., Wagner, K., Soto, D. W., & Baezconde-Garbanati, L. (2009). Parent-child acculturation patterns and substance use among Hispanic adolescents: A longitudinal analysis. *Journal of Primary Prevention*, *30*, 293–313. doi:10.1007/s10935-009-0178-8

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–69.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Zamboanga, B. L., Schwartz, S. J., Jarvis, L. H., & Van Tyne, K. (2009). Acculturation and substance use among Hispanic early adolescents: Investigating the mediating roles of acculturative stress and self-esteem. *Journal of Primary Prevention*, *30*, 315–333. doi:10.1007/s10935-009-0182-z

Child Development, July/August 2013, Volume 84, Number 4, Pages 1373–1391

Associations Between First-Time Expectant Women's Representations of Attachment and Their Physiological Reactivity to Infant Cry

Jennifer C. Ablow, Amy K. Marks, S. Shirley Feldman, and Lynne C. Huffman
University of Oregon

Associations among 53 primiparous women's Adult Attachment Interview classifications (secure–autonomous vs. insecure–dismissing) and physiological and self-reported responses to infant crying were explored. Heart rate, skin conductance levels, and respiratory sinus arrhythmia (RSA) were recorded continuously. In response to the cry, secure–autonomous women demonstrated RSA declines, consistent with approach-oriented responses. Insecure–dismissing women displayed RSA and electrodermal increases, consistent with behavioral inhibition. Furthermore, insecure–dismissing women rated the cries as more aversive than secure–autonomous women. Nine months postpartum, secure–autonomous women, who prenatally manifested an approach-oriented response to the unfamiliar cry stimulus, were observed as more sensitive when responding to their own distressed infant, whereas women classified prenatally as insecure–dismissing were observed as less sensitive with their own infants.

Crying has been identified as one of the most salient attachment behaviors, and the human infant is similar to newborns of other species in its ability to emit vocal signals that serve as a stimulus for what becomes a complex interactional system through which infant and caregiver build their relationship (Broth, Goodman, Hall, & Raynor, 2004; Spangler & Grossmann, 1993). Contemporary biosocial models propose that the attachment relationship is a context in which caregivers play a critical role responding to infant distress and, in the process, regulating infants' behavioral and biological systems (Boyce & Ellis, 2005; Hofer, 2006). As do others (Groh & Roisman, 2009; Mills-Koonce et al., 2007; Moore & Calkins, 2004; Tronick & Cohn, 1989) , we speculate that to respond in an optimal manner that promotes infant organization, parents must first be able to manage their own physiological and emotional response to their infants' cues. The psychobiological reactions to their infants' cries that some parents experience may impact their ability to play a constructive role in their infants' regulation. In addition to questions about what individual difference factors explain how parents respond psychobiologically to child distress, the possibility exists that the bases for maladaptive parental responding may be detectable before the birth of a child. This study examined differences in expectant women's physiological responses to infant cry stimuli based on their own adult representations of attachment. Women were then observed when their infants were 9 months old to investigate whether attachment status and cry reactivity predicted maternal sensitivity in response to their own infant's distress.

Individual Differences in Adults' Response to Attachment Cues

Although being aroused by an infant's cries is typically a stimulus for sensitive maternal responding (mobilization for approach; Ham & Tronick, 2006), infant distress can also be dysregulating, especially when persistent and intense. For some caregivers, infant signals are subjectively aversive, emotionally upsetting (Zeskind & Marshall, 1988), and physiologically arousing (Weisenfeld, Malatesta, Whitman, Granrose, & Vili, 1985). Parents who self-report that they are prone to insensitive and, at times, abusive behavior, respond to infant cry stimuli with greater accelerations in heart rate (HR) and higher rates of respiration than comparison parents (Crowe & Zeskind, 1992; Frodi & Lamb, 1980; Lin, Bugental, Turek, Martorell, & Olster, 2002). When physiologically aroused, these parents appear to respond to their children's actions without reflective appraisal and, instead, draw involuntarily upon relationship schemas that have their origins in their own interpersonal histories (Bugental, Lewis, Lin,

---

Correspondence concerning this article should be addressed to Jennifer C. Ablow, Department of Psychology, 1227, University of Oregon, Eugene, OR 97403-1227. Electronic mail may be sent to jcablow@uoregon.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0019
DOI: 10.1111/cdev.12135

Lyon, & Kopeikin, 2000). Research has shown that when their working models of relationships as adults reflect ongoing insecurity, mothers appear to perceive a narrower range of infant emotions and are typically less accurate in their identification of specific infant emotions than secure–autonomous mothers (Blokland & Goldberg, 1988). Furthermore, research indicates that parents with insecure working models of attachment have greater difficulty interpreting their infants' ambiguous signals (Raval et al., 2001), have longer delays in responding to their infants' signals (Atkinson et al., 2000), and experience more negative affect while responding to their infants (Adam, Gunnar, & Tanaka, 2004) than do parents with secure working models of attachment. As such, a theoretical and empirical basis exists for predicting that differences in adults' attachment representations are associated with variation in how they perceive infants' distress cues.

Perhaps the most widely used and well-validated instrument to assess one's current working model of attachment is the Adult Attachment Interview (AAI; George, Kaplan, & Main, 1996). During the AAI, adults are asked a series of questions regarding their childhood experiences, providing memories relevant to loss, separation, rejection, and trauma. On the basis of the coherence of adults' verbal accounts, AAI interviews are classified as secure–autonomous, insecure–dismissing, or insecure–preoccupied and are thought to represent organized motivational tendencies toward attachment-related information (Main, 2000). In this study, we were specifically interested in women classified as secure–autonomous and insecure–dismissing. Not only are these two classifications the most representative of adult samples (Main & Goldwyn, 1998), but prior research has provided clear evidence that these two classifications, assessed both prenatally and postnatally, have implications for adults' capacity to develop sensitive and responsive relationships with their children (see meta-analysis by van IJzendoorn, 1995). Additionally, recent insights from an emerging literature on the psychophysiology of adult attachment has begun to identify physiological differences in adults' responses to attachment-related challenges, including responses to infant cry (e.g., Dozier & Kobak, 1992; Roisman, 2007; Roisman, Tsai, & Chiang, 2004), as a function of coherence within the AAI.

Psychobiological Responses to Attachment Cues

On the basis of theory originally advanced by Gray (1975), Fowles (1988) outlined evidence for the existence of two antagonistic components of the biological motivational system. The first, known as the behavioral inhibition system (BIS), is considered the anxiety system as it is involved in the effort to inhibit behavioral response when aversive consequences are anticipated. The second system, known as the behavioral activation system (BAS), is involved in approach-related activation. Fowles argued that behavioral inhibition is associated with electrodermal activity, specifically, increased skin conductance levels (SCL), and behavioral activation is associated with cardiovascular reactivity, typically in the form of increased HR.

Recent evidence from the attachment literature partially corroborates Fowles's (1988) hypothesis. For example, in separate studies, Dozier and Kobak (1992) and Roisman and colleagues (Dias, Soares, Klein, Cunha, & Roisman, 2011; Roisman et al., 2004) found that adults who used deactivating strategies while responding to questions during the AAI showed significant electrodermal increases from baseline. Social psychologists identify electrodermal response as a useful indicator of interpersonal processes (Dawson, Schell, & Fillion, 2000), often associated with states of anxiety and efforts to avoid difficult conversational material (Gottman & Levenson, 1992). For example, Roisman and colleagues found that both married (Roisman, 2007) and unmarried (Holland & Roisman, 2010) insecure adults who demonstrated a tendency to use deactivating strategies during the AAI manifested significantly elevated electrodermal responses while discussing areas of disagreement with their marital or romantic partner. Of particular relevance to this study, Groh and Roisman (2009) examined links between college students' secure base knowledge and changes in SCL following exposure to recorded infant cry. As predicted, individuals who generated narratives that were lower in secure base knowledge showed more electrodermal reactivity than did those high on secure base knowledge. In contrast, higher secure base knowledge was associated with little or no change in SCL in response to infant cry over baseline values (Groh & Roisman, 2009). Interestingly, no association was found between secure base knowledge and SCL response to recorded infant laughter, suggesting that BIS responses may be specific to attachment stimuli perceived as aversive or stressful. That Groh and Roisman (2009) identified attachment-linked differences in physiological reactions to infant cry in college students suggests that individual differences in cry reactivity may be in place before adults have infants of their own (Bugental et al., 2000).

Support for the BAS within the attachment literature, primarily examined using HR and HR variability, has been mixed. In their study of adopted adolescents, Beijersbergen, Bakermans-Kranenburg, van IJzendoorn, and Jeffer (2008) report that, relative to secure participants, dismissing adolescents demonstrated less cardiac reactivity as they discussed early attachment experiences during the AAI, yet elevated reactivity during a conflict-solving task with their mothers. In contrast, in the Roisman (2007) study, adults using a deactivating strategy during the AAI did not demonstrate elevated HR reactivity (compared to hyperactivating adults) during a conflict-solving task with their romantic partner. Other studies examining adult attachment strategies and HR variability found no significant differences in response to the AAI (Dias et al., 2011; Roisman et al., 2004) or an attachment-related challenge (Groh & Roisman, 2009). Conflicting and null findings may be because the heart is innervated by both the sympathetic (SNS) and parasympathetic (PNS) branches of the nervous system, which often have antagonistic effects during the response to stress (Cacioppo & Berntson, 1999). Activation of the SNS typically leads to elevations in HR and increased metabolic output typical of "flight-fight" responding. The PNS, by contrast, is responsible for the maintenance of homeostatic functioning and poststress metabolic recovery. However, withdrawal of PNS can enable more rapid and efficient change in HR (Berntson, Norman, Hawkley, & Cacioppo, 2008). Thus, a more accurate interpretation of the motivational response to particular stressors may be had when multiple physiological indicators are studied in response to the same stimulus (Beauchaine, 2001).

The measurement of cardiac vagal tone as indexed by respiration sinus arrhythmia (RSA) has become a common way to investigate PNS effects on self-regulation, information processing, and emotional functioning. RSA, the rhythmic oscillation in heart period accompanying breathing, is a noninvasive measure of cardiac vagal control (Rottenberg, Wilhelm, Gross, & Gotlib, 2002). Researchers have found that RSA is sensitive to environmental demands or challenge, and may serve as an index of underlying regulatory abilities in mammals (Porges, 1995). According to the Porges polyvagal theory (2007), individuals should show modest decreases in RSA (e.g., the removal of the "vagal brake") during challenging situations, removing inhibitory effects on HR and, as such, enabling increases in metabolic output needed for behavioral activity. However, when environmental demands have ceased, RSA increases (e.g., application of the "vagal brake") to promote decreases in HR and a return to a homeostatic or baseline state. Conversely, continued application or increases in vagal control when a behavioral response is indicated, or the failure to release vagal inhibition when a behavioral demand has passed, may reflect less adaptive emotional functioning (Beauchaine, 2001; Bornstein & Suess, 2000; Porges, 2007; Rottenberg et al., 2002).

Few studies have considered adults' parasympathetic response to attachment-related cues. Dias et al. (2011) found evidence that insecure women diagnosed with an eating disorder show poorer vagal tone than secure participants during the AAI. And although Roisman and colleagues (Holland & Roisman, 2010; Roisman, 2007) did collect RSA when studying couples engaged in discussions of interpersonal difficulties, they did not find that RSA differentiated between secure versus insecure or deactivating versus hyperactivating adults, suggesting that more challenging attachment-related experiences are needed to discriminate between secure and insecure adults (Roisman, 2007).

Present Study

The central goal of this study was to investigate associations between expectant women's attachment status and their physiological response to infant cry. We then used these prenatal associations to predict maternal sensitivity to their own infant's distress 9 months postpartum. In the present investigation, participants were presented with two different cry stimuli, which we labeled as the "simple" and "complex" cry conditions. In our simple cry condition, women saw a static video image of a crib while they heard but could not see an infant cry for 2 min. Audio-only presentations of infant cry have been the predominant paradigm in the infant cry literature, but have also raised questions about their ecological validity (Zeskind & Lester, 2001). Accordingly, our complex cry stimulus provided a more typical parenting context where woman could reasonably interpret the infant's cry as distressed. Specifically, in our complex cry condition, women watched 3 min of videotape of a mother and infant playing contentedly (1 min), followed by the infant becoming distressed once the mother departed (1 min.). The complex cry condition ended with a pseudo reunion, during which a female actor, dressed to look like the mother, tried unsuccessfully to soothe the baby (1 min.).

Based on theory reviewed above, and evidence that secure and insecure–dismissing adults show different physiologic responses to attachment cues, we expected that women with secure–autonomous models of attachment would show autonomic reactions consistent with approach-related behavior, in particular, increased HR and decreased RSA. In contrast, we predicted that expectant women with insecure–dismissing attachment models would show an autonomic response consistent with behavioral inhibition, specifically, increased SCL and increased RSA. Our hypotheses reflect our prediction that exposure to an infant's cry would be differentially arousing for women with secure versus insecure models of attachment.

Following each cry stimulus, women were asked to rate the characteristics of the cry as well as to rate how they would have felt emotionally had they been the distressed infant's mother. Infant cry that is described as distressed typically elicits self-reports of concern that are consistent with the initiation of ameliorative care. In contrast, infant cry that is described as aversive can elicit both nonresponsive and insensitive parenting responses (Crowe & Zeskind, 1992). Accordingly, we predicted that women would rate the complex cry more negatively than the simple cry, in terms of both the cry's characteristics (e.g., distressing, urgent, aversive) and how the stimuli would have made them feel had they been the infant's mother (e.g., anxious–overwhelmed, angry–rejected). In addition, we expected self-reported reactions to the cry stimuli would differ according to women's AAI classification. We predicted that insecure–dismissing women would rate both cry conditions as more aversive than secure–autonomous women; secure–autonomous women would rate infant crying as more distressing. We also predicted that secure women, imagining they were the crying infant's mother, would indicate greater feelings of distress, whereas insecure–dismissing women would report greater feelings of anger–rejection. In general, we expected stronger effects for the complex cry condition.

Next, we examined whether prenatal differences in attachment classification and autonomic reactivity to cry predicted women's sensitivity when responding to their own 9-month-old babies' distress following a disruption in social interaction (i.e., a brief separation). We expected that secure–autonomous women would be more sensitive with their infants than insecure-dismissing women. Consistent with Porges (2007), we also hypothesized that women who manifested autonomic patterns in response to infant cry that were consistent with behavioral inhibition would be our least sensitive mothers when responding to their infants' distress; by contrast, women showing patterns of physiological engagement in response to infant cry would show greatest sensitivity with their infants.

In a final set of multivariate regression models, we examined how attachment and physiology might work together to predict maternal sensitivity to infant distress. Although bivariate associations have been identified previously in the literature, the links among adult attachment classifications, physiological response to infant distress, and maternal sensitivity have not yet been well explored (Fox & Hane, 2008; Thompson, 1997). In line with our notion that physiological response to infant distress might shape how parents with different attachment histories respond, we expected that women's prenatal autonomic response to infant cry would mediate the link between attachment and maternal sensitivity. However, such a model presumes that attachment representations temporally precede how women react physiologically to infant cry. Because establishing temporal precedence is needed to validate causal chains (Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001), the conclusions drawn from this study will need to be statistical in nature until experimental or true prospective longitudinal designs can clarify this question. As such, we tested for mediation and moderation by employing the analytic criteria outlined by Baron and Kenny (1986) and refined by Kraemer et al. (2001).

Method

Participants

Sixty-nine primiparous women in their third trimester of pregnancy were successfully recruited through local childbirth education classes, obstetric practices, and hospitals in the mid San Francisco Bay area. Recruitment material informed that the project was investigating the process by which women who were expecting their first baby prepared for motherhood, both psychologically and biologically. The participants ranged in age from 19 to 41 years (M = 32.65 years, SD = 4.38 years). Postnatally, mother and infant dyads were visited in their homes approximately 9 months after the baby's birth; infants were on average 283 days of age (SD = 13 days) during this visit. The majority of the sample was self-identified as European American (77%), with 9% Asian, 5% African American, and 3% of Hispanic origin. The socioeconomic status of the sample was mostly middle to upper-middle class with the average annual household income of $120,000 ($SD$ = $16,500), ranging from less than

$10,000 to more than $200,000. All participants, except one, were married at entry into the project, with average length of marriage of 3.43 years (*SD* = 2.2 years), ranging from 6 months to 10 years. Fifty-four percent of the women had college degrees, 33% had postgraduate degrees, and less than 2% did not have a high school degree. Although no women in the project had given birth previously, 29% of the sample had at least one prior pregnancy.

Procedure

Women visited our research laboratory at approximately Week 32 of gestation. One week prior to the visit, participants received a packet of informed consent documents and research questionnaires by mail. This packet assessed an array of issues, from health history and practices, mental health history, family functioning, current psychosocial functioning, and current symptomatology. Upon arrival at the laboratory, informed consent was documented. Women were asked to sit in the data collection chair facing a TV monitor while a female laboratory assistant placed surface sensors and a respiratory bellow to measure continuously cardiovascular activity, SCL, and respiration. Baseline values were established before women participated in the infant cry stimuli tasks followed by the AAI.

Approximately 9 months postpartum, women were visited in their homes where they completed additional questionnaires and participated in a mother–infant interaction that comprised a free play and an unexpected social disruption. Our 9-month interaction task was designed to bridge the social challenge of the still-face paradigm (Tronick & Cohn, 1989) and the strange situation (Ainsworth, Blehar, Water, & Wall, 1978) by stressing the infant with a brief period of parental unavailability and separation. The purpose of the interaction was to observe maternal behavior in response to infant distress. Although free play has been the dominant mode of assessing maternal sensitivity early in development, research has revealed that observations during distress rather than nonstress contexts (Atkinson et al., 2005) may be more indicative of individual differences in maternal sensitivity that are predictive of attachment-related outcomes. Biringen (1990) has documented that although brief maternal separation in the home is a regular and frequent occurrence, it can cause transient infant distress. Within this study, significant increases in infant distress in response to the brief maternal separation point to the internal validity of this procedure (as reported elsewhere in the Results). During our 9-month interaction, infants and mothers were first placed on a blanket on the floor in a comfortable room in their house (typically, living room). Mothers were instructed to interact with their infants for 2 min without toys and then for 2 more min with soft toys provided by the experimenters. At 4 min, mothers were signaled to leave their infant in the room and go to another room. Mothers returned after 90 s or, if the infant became distressed, sooner. For the 2-min reunion, mothers were instructed to respond to their infants' distress so that they could resume "play as usual." The interactions were videotaped and subsequently coded for both infant distress and maternal behavior. In this study, only maternal behavior on reunion with her baby was used in our central analyses given our focus on maternal response to infant distress (vs. maternal behavior while playing).

Infant Cry Stimuli

Women's continuous physiological reactions to infant cry were measured during two separate cry stimulus conditions, specifically, a simple cry stimulus and a complex cry stimulus. Order of condition presentation (baseline–simple cry/baseline–complex cry or baseline–complex cry/baseline–simple cry) was counterbalanced and assigned to women at random. After prerecorded instructions were administered via videotape, women were familiarized with an electronic rating dial that they would use to provide self-reported reactions following each cry stimulus. Before presentation of the first cry condition, participants watched a 1-min video clip of a tranquil seascape scene during which baseline physiology was measured and after which women reported on their mood state using the rating dial. Next, women watched either the complex cry condition or the simple cry condition.

The complex cry condition consisted of three separate 1-min segments. During the play segment, women watched a gender-ambiguous 9-month-old infant playing contently on the floor with its mother; all that could be seen of the mother was her legs in blue jeans and part of her white t-shirt. At the end of the play segment, participants watched the mother (lower torso) get up and leave the room, at which point the baby began to cry; we refer to this as the mother-departs segment. The mother-departs segment started when the on-screen mother leaves the room and the baby is seen crying on the floor for 60 s. As such, this segment consists of the video image and accompanying audio of the baby crying. We refer to the next 60-s segment as the mother-actor segment, during which an adult woman who was a stranger to the baby, yet dressed exactly like the baby's mother, enters the room. Women watched as the mother actor tried unsuccessfully to soothe the baby. The instruction for the complex and simple cry conditions was, "Please watch the following film." For both conditions, women sat in a comfortable chair that was positioned 1.5 m away from a 36-in. television monitor; women listed to the audio through the monitor's internal speakers. The average fundamental frequency (F_0) of the cry was 653 Hz (range = 454–785 Hz). The average sound pressure level of the cry was 69.61 dBA (range = 65–73 dBA), as measured with a Brüel & Kjær (Brüel & Kjær, Sacramento, CA) precision sound level meter type 2235, which was placed where a subject's head would be.

The simple cry condition used the same audio track as the complex cry condition, but superimposed the track on a still image of a crib; the crib had a baby blanket draped over the sides, obscuring views of any baby that might have been (but was not) in the crib. Thus, in contrast to the complex cry condition, which consisted of video and audiotape of a crying infant, the simple cry consisted of only the audio cry. Following the baseline, women listened to the audio track of the crying infant for 120 s, which we divided into two 60-s segments for analytic purposes—Minutes 1 and 2 of the simple cry—because these 2 min corresponded to the baby's cry during the mother-departs segment and mother-actor segment of the complex cry condition. Following the cry episodes, a female interviewer administered the AAI to all participants. Interviewers underwent extensive training (supervised by the first author, who was trained by Mary Main) and followed a standardized, semi-structured interview script.

Measures

Psychological measures. In addition to a demographics questionnaire, which included questions regarding age, ethnicity, marital status, occupational status, household income, number of prior pregnancies, prior child-care experience, and history of mental health care utilization, participants completed the depression and anxiety subscales of the Brief Symptom Inventory (BSI; Derogatis, 1975). The BSI is a 53-item self-report inventory designed to assess the symptom patterns of respondents. Items are rated on a 5-point scale, ranging from 0 (*not at all distressed*) to 4 (*extremely distressed*). The BSI's reliability (test–retest $rs > .61$), internal consistency (Cronbach's $\alpha > .86$), and convergent validity are strong (see review by Boulet & Boss, 1991).

Adult Attachment Interview. The AAI (George et al., 1996) is a semistructured, clinical interview designed to provide an assessment of an individual's current state of mind with respect to past caregiver–child attachment-related experiences (Hesse, 2008). Interpretations of the adult attachment categories do not rest on the assumption that they represent veridical accounts of early childhood experience. Rather, transcripts of the interviews are coded by trained raters according to how coherently people recount their past. The individual's strategy during the AAI (e.g., derogating or minimizing of attachment vs. valuing and rendering a balanced, coherent narrative despite positivity or negativity of actual experience) reflects the quality or security of one's current state of mind with respect to attachment (Hesse, 2008).

The AAIs were transcribed verbatim, and identifying information was removed prior to coding. To assess individual differences in attachment, transcripts were coded by a certified AAI coder, trained by Mary Main and Erik Hesse, who had achieved greater than 80% agreement with Main on the official reliability test for both three-group (Ds, F, E) and four-group (Ds, F, E, U) coding. For reliability, a second certified AAI coder scored 20 transcripts. Both coders were blind to all other project data and to each other's AAI scores. Interrater agreement between these two coders was 82% ($k = .70$) for the three-group coding and 75% ($k = .64$) for the four-group coding. Coding resulted in the following classification breakdown: 39 (56.5%) secure–autonomous; 17 (24.6%) insecure–dismissing of attachment; 3 (4.3%) insecure–preoccupied; 7 (10.1%) unresolved with respect to trauma; and 3 (4.6%) could not be classified. We focused only on the women classified as secure–autonomous and insecure–dismissing.

Physiological data acquisition and reduction. Measures of physiology were sampled continuously. Two silver–silver chloride electrodes were placed axially on the left and the right rib cage, approximately 10–15 cm below the armpits. A ground electrode was placed on the collarbone. Respiration was collected with pneumatic bellows, placed around the participant's chest and fixed with a metal bead chain. Data were collected with equipment from James Long Company (James Long Co., Caroga Lake, NY).

Heart rate. For electrocardiography (ECG), the bioamplifier was set for band-pass filtering with half-power cutoff frequencies of 0.1 and 1000 Hz, and the signal was amplified with a gain of 500. The pneumatic bellows was attached to a pressure transducer with a band pass of direct current to 4000 Hz so that no segment or time shifts were introduced in the measurement of respiration. Data were digitized at a sampling rate of 1000 Hz with a 12-bit analog-to-digital board in laboratory computer.

R waves were identified in the ECG using an automated algorithm. Although rarely necessary, an interactive graphical program enabled manual correction of missed or misidentified *R* waves. The *R*-wave times were converted into interbeat intervals (IBIs) and resampled into equal time intervals of 125 ms. The prorated IBI values were saved for analyses of the mean and variance of heart period as well as further processing of heart period variability due to RSA.

Respiratory sinus arrhythmia. Respiratory sinus arrhythmia (RSA) represents the rhythmic fluctuation in heart period during cycles of respiration (Grossman & Kollai, 1993; Porges, 2007). In this study, RSA was measured in seconds and computed with the peak–valley method, which has been shown to correlate highly with spectrally derived measures of RSA (Galles, Miller, Cohn, & Fox, 2002). The respiration signal was used to identify the onset of inspiration and expiration times and relative tidal volume. RSA was computed as the difference in IBI between the onset of inspiration and the onset of expiration for consecutive respiratory cycles. As such, RSA was automatically computed twice per breath, once from inspiration to expiration and again from expiration to inspiration. The breath-by-breath differences across respiratory cycles were then averaged so that any underlying trends in heart period had minimal impact on the computed mean RSA.

Skin conductance levels. Participant's SCL (expressed in microSiemens) was examined using two Ag–AgCl skin conductance electrodes filled with Bio-Gel, an isotonic NaCl electrolyte gel. Velcro bands secured electrodes to the volar surfaces of distal phalanges of participant's first and second fingers on their nondominant hand. Double-sided adhesive collars with a 1-cm hole in the center were used to control gel contact. A constant sinusoidal (AC) voltage was used to avoid biasing the electrodes. The voltage was 0.5 V RMS. SCL responses were assessed at a rate of 1,000 readings per second, and SCL was calculated using the James Long Company software. A 14-Channel A/D converter was used to digitize and amplify signals. Averages for SCL responses were obtained for each baseline and segment of the cry conditions.

Graphs were made of each participant's HR, SCL, and RSA across all segments. Data outliers for each participant were examined graphically and statistically. Points above or below 3 *SD* of the mean were removed and an entire case was removed if more than 15% of their data files needed editing. Segment averages for all three physiological measures were computed for the discrete segments in each cry stimuli. For the simple cry episode, averages were computed for the baseline segment, Minute 1 segment, and Minute 2 segment. For the complex cry episode, averages were computed for the baseline segment, infant–mother play segment (1 min), mother-departs cry segment (1 min), and mother-actor cry segment (1 min).

Missing physiological data. None of the women from the overall sample of 69 was excluded because their data files required more than 15% editing. However, 4 of 69 women (5.7%)—2 secure–autonomous, 1 insecure–dismissing, and 1 women classified as unresolved with respect to trauma—had to be treated as missing because our equipment failed to detect accurate physiological levels. Consequently, in this study, our sample (*N* = 53) comprised 37 women classified as secure–autonomous and 16 women classified as insecure–dismissing.

Self-Report Ratings at Baseline, of Cry and Feelings as Mother

Using procedures and scales developed by Gross and Levenson (1995), women used a rating dial to describe their mood state before the cry tasks began. Using the dial with a scale of 1 (*not at all*) to 7 (*very much*), in response to a video-recorded prompt, women indicated how they felt on three dimensions: anxious, happy, and relaxed. Following each cry stimulus, five rating scales developed by Zeskind and Lester (1978) were used to quantify women's reports of how unpleasant or aversive they perceived the cries. Using a 7-point scale, women rated the quality of the infant's cry on five cry dimensions: urgent, soothing, piercing, grating, and distressed. The polarity of the scales was consistent in the scoring (e.g., the highest level of grating was represented by 7 and the lowest by 1). High intercorrelations ($rs > .68$) and a principal components analysis (all eigenvalues > 1.92) of these scales yielded the following reduced set of cry scales: urgent (urgent and soothing, reversed), aversive (grating and piercing), and distressed. Using the same rating dial, participants were asked to indicate how they would have felt had they been the crying infant's mother on eight scales that were developed for this study, but patterned after the Description of the Child's Behavior Questionnaire (Zeanah et al., 1993), which asks participants to infer the emotional state of a distressed infant. Asked to infer how they would have felt if they had been the distressed baby's mother, participants rated the following eight dimensions: angry, rejected, anxious, overwhelmed, loved, needed, happy, and patient. The polarity of the 7-point scales was consistent in the scoring (e.g., the highest level of rejected was represented by 7 and the lowest by 1). High intercorrelations ($rs > .71$) and a principal components analysis (all eigenvalues > 1.48) of these scales yielded the following reduced set of cry scales: angry–rejected (angry and rejected), alarmed (overwhelmed and anxious), and loved (loved, needed, happy, and patient).

Maternal Behavior With Infant at 9 Months Postpartum

Maternal sensitivity was assessed following a brief separation from her infant using an adaptation of the global ratings of mother–infant interaction (Murray, Fiori-Cowley, Hooper, & Cooper, 1996). Coders coded five dimensions of maternal behavior on a scale from 1 to 5. Higher scores represented higher levels on five dimensions: warmth, acceptance, responsiveness, demandingness (reversed), and sensitivity. *Warmth* was defined as the degree to which the mother expressed love and affection toward her baby, *acceptance* included the willingness and ability of the mother to follow the infant's lead, *responsiveness* was operationalized as the mother's awareness of her infant's signals and response to them, *demandingness* was defined as the degree to which the mother required the infant to behave in a certain way, and *sensitivity* included mothers to identify her infant's signals and vary behavior appropriately.

A subset of tapes (20; 43%) was coded by two coders to evaluate interrater reliability. Intraclass correlations between coders for each of the five dimensions coded were .82 (warmth), .82 (accepting), .95 (responsive), .87 (demanding, reversed), and .90 (sensitivity). Because the scales were highly intercorrelated (mean $r = .67$, range = .43–.88) and accounted for by a single factor in a principal components analysis, we averaged each woman's score on all five dimensions, creating a single measure of maternal sensitivity to her infant's distress.

Infant Distress

Infant distress was coded during the initial minute of play and during the entire 90-s separation phase of our infant–mother interaction to determine which infants became distressed in response to their mother's unexpected departure. Following a procedure similar to Haley and Stansbury (2003), distress was coded in 1-s intervals; for each second of interaction, infants received a distress score (1 point) if they protested, cried, or fussed, or 0 point for nondistressed behavior. Percentage scores of infant distress were calculated by summing the total number of seconds in each episode that distress occurred. This sum was then divided by the total interaction time per episode. Coders reliably observed infant distress (percent agreement = .84).

Results

Preliminary Analyses

Preliminary analyses were conducted to determine whether key health, mental health, and psychosocial variables were associated with our primary variables of interest. As shown in Table 1, AAI classification was unrelated to all demographics except for average level of educational attainment, $F(1, 52) = 7.23$, $p < .01$, with women classified as secure typically having completed college and women classified as insecure–dismissing typically having completed some though not all college. None of these demographic variables was associated significantly with women's physiology or with maternal sensitivity postpartum. Because inclusion of education level in our central analyses did not alter our findings when tested as main or interactive effects, these variables were dropped and are not discussed further.

Table 1 presents women's self-reported mood ratings and their physiological levels of arousal during the baseline segment prior to exposure to an initial cry condition. A multivariate analysis of variance (MANOVA) conducted on women's baseline mood ratings indicated that women classified as secure–autonomous and insecure–dismissing were statistically equivalent during our pretask baseline, $F(3, 49) = 1.31$, $p = .28$. MANOVAs conducted on women's physiological baselines also revealed no significant differences in levels of autonomic arousal between secure–autonomous and insecure–dismissing women, $F(3, 49) = 1.93–2.17$, $p > .12$. This general pattern of self-report and physiological equivalence at baseline increased confidence that subsequent divergences between secure–autonomous and insecure–dismissing women might reasonably be attributed to differences in women's response to our cry stimuli (Wainer, 1991).

Table 1
Demographic and Baseline Differences Between Women Classified as Autonomous or Dismissing on the Adult Attachment Interview (AAI)

	Total sample (N = 53)		AAI autonomous (n = 37)		AAI dismissing (n = 16)	
	M (SD) or % (n)		M (SD) or % (n)		M (SD) or % (n)	
Demographics						
Age	31.91	(4.17)	32.16	(3.56)	31.55	(5.33)
Household income	\$119,000	(\$9K)	\$121,000	(\$10K)	\$116,000	(\$9K)
Length of marriage in years	3.11	(1.85)	3.22	(1.71)	3.10	(2.24)
Education[a]	4.11	(.79)	**4.35**	**(.54)**	3.81	(.91)
Race/ethnicity						
African American	1.9%	(1)	0%	(0)	6.3%	(1)
European American	78.8%	(41)	83.8%	(31)	68.8%	(11)
Latina	3.8%	(2)	2.7%	(1)	6.3%	(1)
Asian	9.6%	(5)	10.8%	(4)	6.3%	(1)
Other	5.8%	(3)	2.7%	(1)	12.5%	(2)
Previous pregnancies ("yes")	26.9%	(11)	29.7%	(11)	18.8%	(3)
Prior child-care responsibilities ("no" = 1, "yes" = 2)	1.23	(.19)	1.25	(.23)	1.21	(.19)
History of intrapartum medical problems ("no")	92.4%	(49)	91.9%	(34)	93.8%	(15)
History of psychological counseling ("yes")	38.3%	(21)	43.2%	(16)	31.3%	(5)
BSI depressive symptom subscale	0.24	(.29)	0.25	(.37)	0.22	(.26)
BSI anxiety symptom subscale	0.34	(.30)	0.36	(.41)	0.33	(.24)
Baseline measures						
Anxiousness	1.54	(1.09)	1.42	(.76)	1.67	(1.37)
Happiness	4.80	(1.18)	4.88	(.99)	4.71	(1.36)
Relaxed	5.44	(1.45)	5.35	(1.41)	5.54	(1.50)
Heart rate	87.92	(10.24)	89.22	(10.84)	85.95	(9.10)
Skin conductance level	9.77	(4.45)	9.62	(4.10)	10.12	(5.21)
RSA	.045	(.02)	0.040	(.02)	0.055	(.02)

Note. Baseline physiology and self-reported rating data were derived from the resting period before the first cry stimulus task. Bold mean (standard deviation) indicated that one AAI classification's value was significantly different from the other AAI classification.
[a]Education coding: 3 = some college, 4 = 4-year college graduate.

The order in which the cry stimuli were presented was evaluated to see if it affected women's physiological responses. Although presentation order was significant in only 7 of 42 (16.7%) effects tested, we covaried potential order effects in our analyses but do not comment on it further.

Women's Self-Reported Reactions to the Cry Stimulus

As a way to understand how women perceived our cry stimuli, we first analyzed their self-reported reactions to our cry stimuli using the rating dial data. A repeated measures MANOVA was used to evaluate women's ratings of the characteristics of the infant cry (cry scales) and in a second model, how they would have felt had they been the baby's mother (maternal scales). Significant multivariate effects were found for the different cry scales, $F(3, 45) = 38.91$, $p < .000$, $\eta_p^2 = .36$; cry condition, $F(1, 50) = 3.94$, $p = .045$, $\eta_p^2 = .21$; and a significant AAI × Cry Scale interaction, $F(1, 50) = 3.83$, $p = .053$, $\eta_p^2 = .19$. Averaged across both cry conditions, women rated the cry stimuli as significantly more urgent ($M = 5.39$, $SD = 0.98$) than they did distressed ($M = 4.30$, $SD = 1.38$), and significantly more distressed than they did aversive ($M = 3.57$, $SD = 0.98$). These results indicate that both cry conditions created a reasonably salient stimulus. Collapsing across scales, women rated the qualities of the complex cry condition ($M = 4.65$, $SD = 1.73$) more negatively than the qualities of the simple cry condition ($M = 4.02$, $SD = 1.48$). A probe of the significant interaction revealed that insecure–dismissing women rated both the simple and complex cry conditions as more aversive than secure–autonomous women, $t(51) = 3.09$ and 3.29, $p < .05$, respectively. Contrary to prediction, there were no significant differences between secure–autonomous and insecure–dismissing women's rating of the cries' urgency or level of distress.

Analysis of the maternal scales yielded a significant multivariate effect, $F(2, 49) = 18.97$, $p < .000$, $\eta_p^2 = .25$, and a significant three-way interaction between Maternal Scale × Cry Condition × AAI

Status, $F(2, 49) = 5.71$, $p = .021$, $\eta_p^2 = .19$. Our probe of the interaction revealed that secure–autonomous and insecure–dismissing women rated the three maternal scales similarly following the simple cry condition. However, compared to secure–autonomous women ($M = 1.37$, $SD = 0.87$), insecure–dismissing women ($M = 1.47$, $SD = 1.57$) reported significantly higher levels of anger–rejection, $t(51) = 2.73$, $p = .018$, and significantly lower levels on the alarmed scale, $t(51) = 2.31$, $p = .023$ ($Ms = 4.15$ and 3.23, $SDs = 1.16$ and 1.41, respectively), in response to the complex cry condition. There were no significant differences between secure and insecure women's ratings on the loved–needed scale.

Physiological Responses to Cry Stimulus

Tests of the study's central physiological questions were conducted using separate repeated measures MANOVAs with planned contrasts to test-specific differences between and within the cry conditions (means are shown graphically in Figure 1 and results are presented in Table 2). MANOVA is particularly useful for analyzing repeated measures as the variance–covariance matrix of the error terms taking into account all dependencies between repeated measures (Bagiella, Sloan, & Heitjan, 2000). In all models, women's classification as secure–autonomous or insecure–dismissing on the AAI served as the two-level between-subjects factor. The within-subject factor comprised the different segments of each cry condition. The first planned contrast compared (a) women's physiological responses during the cry segment of each condition, specifically, the average of both simple cry minutes against the average of the mother-departs and mother-actor segments of the complex cry condition. The next two contrasts focused on the simple cry condition, specifically, (b) how the baseline and first minute of the simple cry condition differed, and (c) how the first and second minutes of the simple cry condition differed. The next three contrasts focused on the complex cry condition, specifically, differences between (d) the baseline and the mother–infant play segments, (e) the mother–infant play and the mother-departs segments, and (f) the mother-departs and the mother-actor cry segments. Because we held specific hypotheses about these contrasts, alpha was set at .05.

Heart rate. The analyses of HR revealed (Table 2) a nonsignificant effect of AAI classification and a nonsignificant effect of cry segment. Contrary to prediction, the multivariate test of the interaction between women's AAI classification and the cry segments of both conditions was not significant, suggesting that our cry stimuli affected the HR of

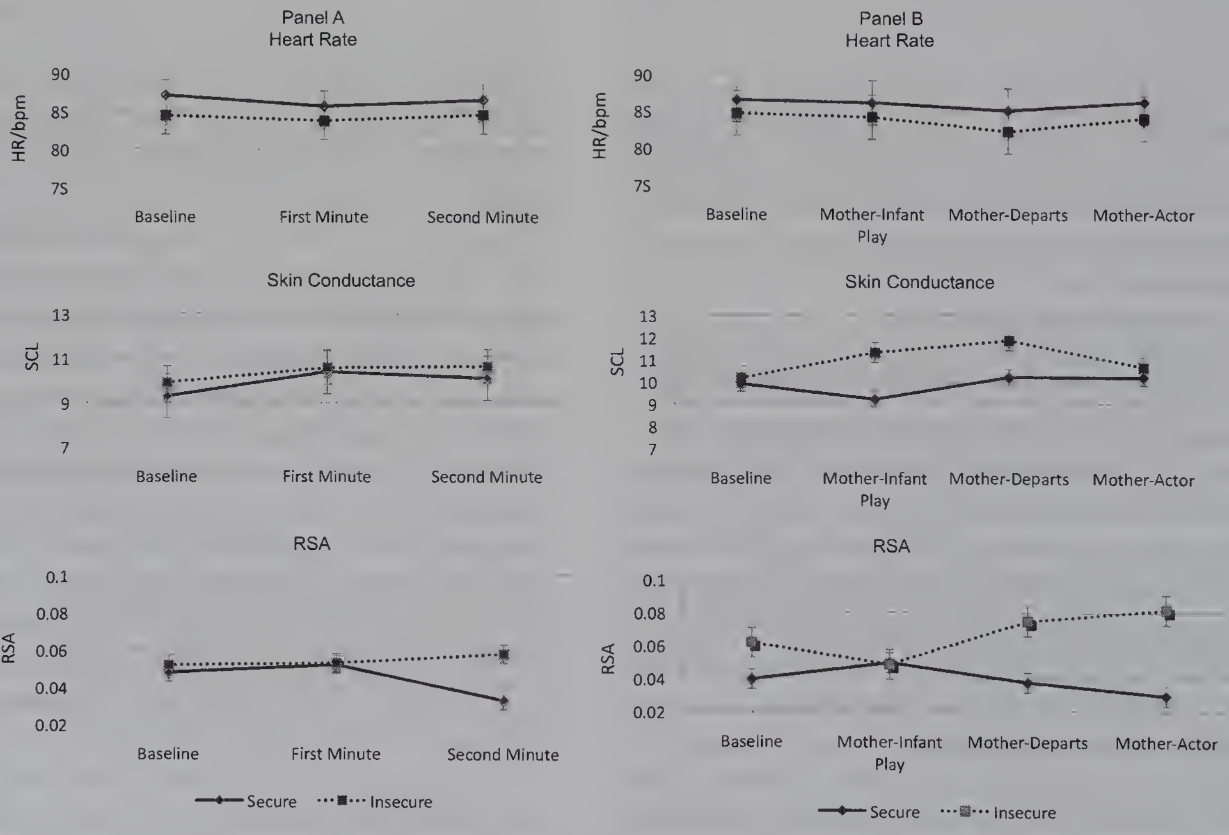


Figure 1. Heart rate (HR), skin conductance level (SCL), and respiratory sinus arrhythmia (RSA) during the simple cry stimulus (Panel A) and complex cry stimulus (Panel B).

Table 2
Multivariate Analysis of Variance (MANOVA) and Planned Contrasts Output for Women's Heart Rate (HR), Skin Conductance Level (SCL), and Respiratory Sinus Arrhythmia (RSA)

MANOVA of main and interaction effects	HR			SCL			RSA		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
AAI classification (df_{num} = 1, df_{den} = 51)	2.47	.075	.054	.097	.917	.001	16.05	.000	.239
Cry segments (df_{num} = 6, df_{den} = 46)	1.40	.160	.103	7.15	.008	.382	.89	.517	.103
AAI × Cry Segments (df_{num} = 6, df_{den} = 46)	0.28	.942	.032	3.67	.047	.212	3.15	.011	.291

Analysis of planned contrasts	Effect	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Comparison of cry stress segments	Segment	2.30	.135	.04	4.23	.041	.12	1.09	.301	.00
	AAI × Segment	.31	.579	.01	.16	.694	.00	3.35	.073	.10
Baseline versus first minute of simple cry	Segment	1.28	.264	.01	1.04	.300	.02	.15	.837	.00
	AAI × Segment	.09	.953	.00	1.54	.220	.03	.19	.855	.00
First versus second minutes of simple cry	Segment	2.09	.154	.01	.41	.154	.09	.91	.345	.01
	AAI × Segment	.54	.464	.01	2.23	.142	.04	5.67	.018	.19
Baseline versus mother–infant play during complex cry	Segment	4.23	.046	.09	3.31	.078	.09	.01	.943	.00
	AAI × Segment	.23	.631	.01	6.75	.012	.20	1.05	.310	.01
Mother–infant play versus mother departs during complex cry	Segment	0.77	.144	.03	4.08	.036	.17	1.09	.302	.01
	AAI × Segment	.50	.483	.01	.18	.752	.00	7.08	.010	.23
Mother-departs cry versus mother actor during complex cry	Segment	3.21	.079	.05	1.37	.271	.15	.75	.391	.02
	AAI × Segment	.42	.520	.01	4.53	.065	.18	4.17	.046	.11

Note. Each contrast comprised two effects: a main effect for the difference between the contrasted segments and an effect for the interaction between women's AAI status and the contrasted cry segments; each contrast has a corresponding F-test value. AAI = Adult Attachment Interview.

secure–autonomous and insecure–dismissing women similarly.

Skin conductance level. The analysis of SCL produced (Table 2) a nonsignificant effect of AAI classification yet a significant effect for cry condition; collapsed across AAI classifications, women's SCL during the complex cry condition was significantly higher than during the simple cry condition, suggesting that the complex cry condition elicited a larger electrodermal response than the simple cry condition. This effect was subsumed, however, by the significant interaction between cry segments and women's AAI classification. Results from the planned contrasts provide a more complete picture of the nature of this interaction.

As shown in Table 2, none of the simple cry condition contrasts for SCL reached significance. Secure–autonomous women and insecure–dismissing women did, however, respond to the complex cry condition differently. Contrast 4 was significant and when probed revealed that secure–autonomous women showed a marginally significant decline in their mean SCL between baseline and the mother–infant play segment, $t(36) = 2.00$ $p = .065$, whereas insecure–dismissing women showed a significant increase in mean SCL from baseline to play, $t(15) = -2.29$ $p = .042$. Contrast 5 revealed that both AAI groups responded to the mother-departs segment with a significant increase in SCL. It was unexpected that both groups responded to the mother-departs segment with increased electrodermal activity. However, an additional t test revealed that whereas the insecure–dismissing women showed a clear rise in SCL between baseline and the mother-departs segment, $t(15) = -5.61$ $p = .009$, there was no difference between the baseline SCL and mother-departs SCL for the secure–autonomous women, $t(36) = .92$ $p = .392$; this may suggest that secure–autonomous women were returning to baseline SCL, a possibility that seemed further supported by the interaction in Contrast 6. Specifically, between the mother-departs segment and the final mother-actor segment, secure–autonomous women's SCL remained statistically unchanged, $t(36) = 1.06$, $p = .229$, whereas the SCL of insecure–dismissing women declined marginally, $t(15) = 2.27$, $p = .062$.

Respiratory sinus arrhythmia. The analysis of RSA revealed (Table 2) a significant effect of AAI classification, a nonsignificant effect of cry segments, and, as predicted, a significant interaction between cry segments and AAI classification. Across all cry segments, secure–autonomous women had lower mean RSA levels ($M = .04$, $SD = 0.02$) than insecure–dismissing women ($M = .07$, $SD = 0.04$).

As shown by Contrast 2, there was no change in RSA for either AAI group from baseline to the first minute of the simple cry condition. However, there was a significant difference between the first and second minutes of the simple cry (Contrast 3), with secure–autonomous women showing a significant decline in RSA, $t(36) = 3.35$, $p = .002$, and insecure–dismissing women showing no significant change in RSA, $t(15) = -1.12$, $p = .281$.

Women's RSA levels did not change significantly from the complex cry baseline to the mother–infant play segment (Contrast 4). The two AAI groups did, nonetheless, respond differently to the subsequent aspects of the complex cry condition. Contrast 5 was significant and when probed revealed that secure–autonomous women showed a significant decline in their mean RSA from the mother–infant play and mother-departs segments, $t(36) = 2.56$, $p = .014$, whereas insecure–dismissing women showed a significant increase, $t(15) = -3.51$ $p = .025$. Contrast 6 also was significant, and showed that the RSA of secure–autonomous women continued to decrease albeit not significantly between the mother-departs and the mother-actor segment, $t(36) = 1.85$, $p = .072$, whereas insecure–dismissing women showed a significant increase during the mother-actor segment, $t(15) = -2.30$ $p = .040$.

Predictions of Maternal Sensitivity at 9 Months Postpartum

In a final set of hierarchical multiple regression analyses, we explored longitudinal associations between women's attachment representations, prenatal physiological reactions to our cry stimuli, and postnatal observations of women's sensitivity with their 9-month-old infant following a brief separation designed to cause the infant distress. The 1-year gap between women's prenatal and postnatal visit to our laboratory resulted in the attrition of 6 (11%) subjects (5 secure-autonomous and 1 insecure-dismissing); as such, our sample size for these analyses was $n = 47$ (32 secure and 15 insecure) women. To limit the number of models used to predict maternal sensitivity, we only utilized women's physiological responses during the acute cry segment of each cry condition (i.e., Minute 2 of the simple cry and the mother-departs segment of the complex cry). Infant distress in response to the separation was examined first.

A 2 (infant distress while playing with mother and during the separation) × 2 (AAI classification) repeated measures ANOVA revealed that the infants' frequency of distress increased significantly between the play segment ($Ms = 5.77$, $SD = 9.22$) and the separation segment of the mother–infant interaction ($Ms = 25.91$, $SD = 20.18$), $F(1, 45) = 4.68$, $p = .036$, $\eta_p^2 = .10$. However, a nonsignificant AAI, $F(1, 45) = 1.94$, $p = .35$, $\eta_p^2 = .01$, and AAI × Task Segment interaction, $F(1, 45) = 0.79$, $p = .72$, $\eta_p^2 = .00$, meant that the infants in both AAI groups were similarly distressed.

Higher infant distress was not significantly associated with maternal AAI classification ($r = .17$, *ns*), but it was associated with greater maternal sensitivity ($r = .29$, $p = .03$). In terms of women's physiology, infant distress was not associated with mothers' prenatal HR response to cry ($rs = .18$ and $.14$, $p > 1.0$, simple and complex cry, respectively) or with SCL ($rs = -.18$ and -20, $p > .10$, simple and complex cry, respectively), but was significantly associated with RSA ($rs = -.29$ and $-.35$, $p < .05$, simple and complex cry, respectively); greater distress was associated with lower levels of RSA during Minute 2 of the simple cry and during the mother-departs segment of the complex cry. Because infant distress in response to our mother–infant separation task showed associations with both predictor and outcome variables, we controlled for infant distress in our multivariate regression models.

Based on eligibility criteria outlined by Kraemer et al. (2001), the results associated with Table 2 constrained the models we could actually test. Specifically, our inability to detect significant bivariate associations between AAI and women's HR and SCL response to the cry stimuli precluded tests of mediation but not moderation. Alternatively, a significant bivariate association between AAI and RSA enabled us to test for mediation, but not moderation. With women's sensitivity as criterion, infant distress was entered first to control for individual differences in infants' response to our separation task. As shown in Table 3, we ran separate models for each physiological system as well as separately for each cry stimulus. In all models tested, AAI classification was a significant independent predictor of maternal sensitivity ($ps < .05$), above and beyond the contributions of infant distress and prenatal physiology. As predicted, women's state of mind with regard to attachment predicted individual differences in their sensitivity to their own infant's

Table 3

Prenatal Predictors of Postnatal Maternal Sensitivity—Tests of Mediation and Moderation

	Maternal sensitivity							
	Simple cry condition[a]				Complex cry condition[b]			
Predictors	β	*B*	95% CI	% var. exp.	β	*B*	95% CI	% var. exp.
HR models—moderation								
Infant distress	.19†	.38	−.30 to .47	.06	.18†	.35	−.32 to .49	.05
AAI	−.34*	−.62	−1.04 to −.19	.11	−.33*	−.60	−1.10 to −.21	.12
HR	.67	.05	−.02 to .14	.02	.72	.09	−.03 to .11	.03
AAI × HR	.50	.02	−.02 to .08	.01	.53	.06	− 01 to .07	.02
	$R^2 = .20$, $F(4, 43) = 2.91$, $p = .042$				$R^2 = .22$, $F(4, 43) = 3.67$, $p = .021$			
SCL models—moderation								
Infant distress	.15	.36	−.32 to 1.05	.04	.17†	.36	−.31 to 1.00	.05
AAI	−.32*	−.59	−1.03 to −.17	.12	−.34*	−.61	−1.05 to −.16	.13
SCL	−.34	−.06	−.51 to .10	.01	−.33	−.05	−.15 to .08	.01
AAI × SCL	−.48†	−.11	−.15 to .06	.06	−.52*	−.20	−.29 to −.03	.08
	$R^2 = .23$, $F(4, 43) = 4.18$, $p = .013$				$R^2 = .27$, $F(4, 43) = 4.70$, $p = .009$			
RSA models—mediation								
Infant distress	.19†	.37	−.30 to .44	.06	.19†	.37	−.30 to .44	.06
AAI	−.35*	−.63	−1.01 to −.15	.12	−.35*	−.63	−1.01 to −.15	.12
	$R^2 = .18$, $F(2, 45) = 4.67$, $p = .016$				$R^2 = .18$, $F(2, 45) = 4.67$, $p = .016$			
Infant distress	.20†	.39	−.28 to .43	.06	.18†	.37	−.30 to .51	.05
RSA	−.27**	−2.24	−10.42 to −.05	.10	−.32**	−4.07	−12.61 to −.22	.13
	$R^2 = .16$, $F(2, 45) = 3.99$, $p = .025$				$R^2 = .18$, $F(2, 44) = 4.85$, $p = .010$			
Infant distress	.18†	.37	−.32 to .50	.05	.18†	.39	−.30 to .49	.05
AAI	−.31*	−.52	−.97 to −.18	.11	−.30*	−.49	−1.01 to −.20	.12
RSA	−.37**	−5.12	−14.07 to −1.90	.10	−.40**	−5.79	−15.24 to −1.99	.13
	$R^2 = .26$, $F(3, 44) = 5.25$, $p = .007$				$R^2 = .30$, $F(3, 44) = 6.21$, $p = .001$			

Note. AAI = Adult Attachment Interview; HR = heart rate; SCL = skin conductance level; RSA = respiratory sinus arrhythmia; % var. exp. = percent of maternal sensitivity explained by predictor.

[a] Physiological data collected during Minute 2 of the simple cry condition used in the simple cry models.

[b] Physiological data collected during the mother-departs phase of the complex cry condition used in models.

†$p < .10$. *$p < .05$. **$p < .01$.

distress 1 year later, with secure–autonomous women showing significantly more sensitivity to their infants than insecure–dismissing women.

As shown in Table 3, women's prenatal HR responses during the simple and complex cry conditions did not predict sensitivity, either alone or in interaction with AAI. In both SCL models, SCL did not directly predict sensitivity, nor did it interact with AAI in the simple cry model. However, the AAI × SCL interaction was significant in the complex cry model. To probe this interaction we computed simple intercepts and simple slopes to represent the association between attachment status and maternal sensitivity (following Cohen, Cohen, West, & Aiken, 2003) at lower (−1 *SD*) and higher (+1 *SD*) levels of women's SCL. As shown in Figure 2, analyses revealed that the association between attachment status and behavioral sensitivity was significant among women who responded to our complex cry with higher SCL, $B = -.58$, $SE = .20$, $p < .01$, but not among women with low SCL, $B = -0.21$, $SE = .18$, *ns*; insecure–dismissing women who responded with high SCL during the complex cry were significantly less sensitive in response to their distressed infant than secure and insecure–dismissing women with low SCL.

As shown in Table 3, neither model involving RSA yielded statistically significant support for the prediction that prenatal RSA mediated the link between AAI and women's sensitivity in response to their infants' distress. As well, we tested for and found that the AAI × RSA interaction terms were nonsignificant, although this test was technically precluded by criteria used to guide tests of moderation (Kraemer et al., 2001). Importantly, however, in both the simple cry and the complex cry models, both AAI and RSA independently predicted maternal sensitivity in response to infant distress. Specifically, women who were secure–autonomous on the AAI and women who responded with

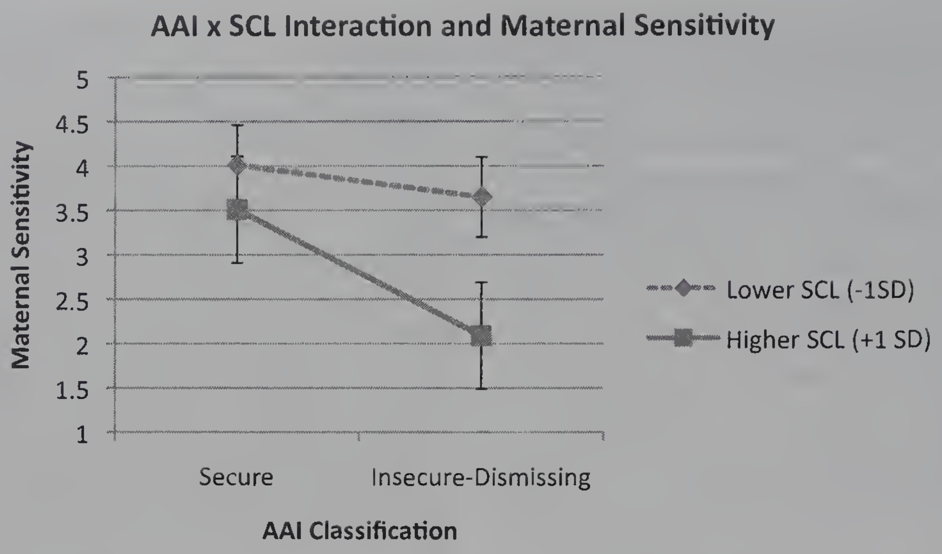


Figure 2. Adult Attachment Interview (AAI) × Skin Conductance Level (SCL) Interaction and maternal sensitivity at 9 months postpartum.

lower levels of RSA to our prenatal cry tasks showed higher levels of maternal sensitivity with their own infant.

Discussion

The results support the hypothesis that women's psychological and physiological responses varied as a function of their attachment classification on the AAI, and prenatal patterns of cry reactivity predicted individual differences in women's maternal sensitivity to their own infants' distress. As in previous studies of cry reactivity (Del Vecchio, Walter, & O'Leary, 2009; Groh & Roisman, 2009; Zeskind & Lester, 2001), the women in this study were presented with levels of infant cry that are typically reached by infants when distressed and that are typically perceived as distressed by parents (Del Vecchio et al., 2009; Frodi & Lamb, 1980). When an infant's cry is perceived as distressed, or engenders a sense of alarm in a mother, women indicate that they would respond more quickly than if they did not perceive the cry as distressed (Wood & Gustafson, 2001). Although secure–autonomous and insecure–dismissing women in this study rated both the simple (audio-only) and complex cry (audio and video) stimuli as comparably distressing, secure women reported feeling more alarmed by the cry, whereas insecure–dismissing women experienced the cry as more aversive. When infant cry is perceived as aversive, some mothers are found to be at risk of withdrawing from their infants (Zeskind & Lester, 2001) or responding too quickly and, potentially, insensitively (Del Vecchio et al., 2009). Furthermore, in response to the complex cry, insecure–dismissing women reported feeling greater anger and rejection than secure women. These self-report data are consistent with evidence that distressed infant cries can elicit orthogonal reactions (Zeskind & Marshall, 1988), although the basis for these differences has not been well identified. These data suggest that secure–autonomous women were concerned by the infant crying, whereas the insecure–dismissing women seemed to experience our cry stimuli as somewhat aversive and angering.

Analyses of our physiological data also revealed that secure–autonomous and insecure–dismissing women manifested reasonably distinct autonomic responses to our cry stimulus. Contrary to prediction, the average HR of both secure–autonomous and insecure–dismissing women remained comparable across both cry conditions. The lack of HR findings in this study is not unparalleled in the attachment literature (Holland & Roisman, 2010; Roisman, 2007). Animal and human studies comparing the effects of reward incentives versus punishment consistently find that physiological reactivity increases in response to appetitive or hedonic motivation (i.e., anticipated reward), but less so in response to increases in aversive motivation (i.e., anticipated punishment; Cacioppo & Berntson, 1999). Although the motivation to soothe a distressed infant may be an important objective for parents, the actual act of listening to a distressed infant may not be experienced physiologically as a reward, per se.

In response to our simple cry condition, secure–autonomous and insecure–dismissing women showed neither significant nor different SCL. This finding is in contrast to Groh and Roisman's (2009) study, in which lower levels of secure base knowledge were associated with heightened SCL from baseline in response to an audio recording of infant crying. It is possible that Groh and Roisman's use of headphones and higher sound pressure levels for their cry stimulus ($M_{\text{dBA}} = 89.51$) affected electrodermal responses more acutely than in our study where women listened to cry via a TV monitor and at lower volumes ($M_{\text{dBA}} = 69.61$). Our participants did manifest significant electrodermal responses during the complex cry condition, which we had designed to be more stressful, as participants could see and hear the distressed infant. On the whole, expectant women classified as insecure–dismissing showed a significant increase in their SCL between our baseline segment and the period during which they heard and saw the infant cry following its mother's departure. This increase may reflect activation of the BIS, or physiological efforts to inhibit behavioral responses when aversive consequences are anticipated (Dozier & Kobak, 1992; Fowles, 1988; Roisman, 2007). Importantly, these data also are consonant with Groh and Roisman's (2009) study that found electrodermal increases to infant crying among college students with limited secure base script knowledge. Bowlby (1980) argued that insecure–dismissing adults passively or actively employ a strategy of deactivating the attachment system to defend against, discount, or ignore attachment cues that would ordinarily increase appetitively motivated behavior. In this study, women classified as dismissing of attachment responded to seeing and hearing a distressed infant's cry in a psychophysiological manner that was potentially consistent with the motivation to avoid the attachment bids of a crying infant.

Insecure–dismissing women also responded with a significant increase in SCL while viewing the infant playing contently with its mother. This is in contrast to the nonsignificant but measurable decrease in the SCL of secure–autonomous women in response to the mother–infant play segment. To the extent that women generally rate positive infant affect as pleasurable (Zeanah et al., 1993), we expected women in our study to show little or no autonomic response to the infant–mother play segment. It is possible that the insecure–dismissing women in our study were not reacting to the infant–mother play condition per se but were instead anticipating an impending stressor. Adults prone to anxious symptoms tend to show elevated sympathetic reactivity and heightened attention to environmental stimuli when negative experiences are expected (Berntson, 2008). However, this electrodermal increase remained meaningful for insecure–dismissing women even when controlling for anxious and depressed symptomatology. Given the tendency to distance themselves from attachment-related material, it is plausible then that these expectant women experienced the mother–infant play epoch as mildly aversive.

In response to the simple cry and paralleling SCL results, insecure–dismissing women showed little RSA response. However, they did show steadily increasing RSA across the complex cry condition. Consistent with our initial predictions, polyvagal theory (Porges, 2007), and paralleling work by Tronick and colleagues (Ham & Tronick, 2006), the increase in RSA exhibited by the insecure–dismissing women may reflect parasympathetic attempts at self-regulation following sympathetic activation, and is generally not indicative of efforts to mobilize infant soothing behaviors. From the standpoint of attachment theory, these insecure–dismissing women may have been engaged in a strategy of regulation designed to exclude or to divert their attention away from an attachment-related stressor. Regulating one's arousal before attending to one's infant may enable some parents to respond more sensitively. However, there is evidence that parents classified as insecure–dismissing tend to respond to their children's attachment bids with delay, inconsistency, or irritation (Main, 2000). In this study, insecure–dismissing women reported that they would have felt more anger and rejection as the crying infant's mother. Although these self-reported reactions to our cry stimuli do not speak conclusively to women's mind-set during our cry conditions, along with the physiological findings they do suggest that insecure–dismissing women may be engaged in inhibiting a behavioral response that is partially related to an emotionally negative response to infant attachment cues.

Secure–autonomous women, by comparison, showed a significant decrease in RSA in response to both cry conditions, a pattern suggestive of the parasympathetic withdrawal needed to initiate a behavioral response (Beauchaine, 2001). The parasympathetic nervous system enables restorative functions, for example, slowing HR. However, the efferent fibers to the heart that originate in the nucleus ambiguous also enable increased metabolic output when a behavioral response is seen as adaptive or when needing to respond to an emotional stressor (Beauchaine, Gatzke-Kopp, & Mead, 2007). Studies linking attachment representations to parenting behavior demonstrate that adults classified as secure–autonomous typically tend to respond to their infant's bids in an attuned, sensitive, and contingent manner (Crowell & Feldman, 1988; Ward & Carlson, 1995). The RSA reduction shown by the secure–autonomous women in response to the infant's distress is consistent with behavioral. Moreover, such an interpretation is consonant with Porges (2007) and parallels evidence that emotionally sensitive mothers show RSA decreases when finally able to respond to their infant after the neutral segment of the still-face paradigm (Ham & Tronick, 2006; Moore et al., 2009) or during the reunion phase of the strange situation procedure (Hill-Soderlund et al., 2008).

Our final goal was to link prenatal attachment status and ANS cry response to women's response to their own, distressed infant. Our finding that women's prenatally assessed AAI status predicted maternal sensitivity almost a year later adds to the corpus of evidence linking adult attachment representations to their own behavior as mothers (van IJzendoorn, 1995). Particularly noteworthy is that this association was between prenatal AAI status and maternal sensitivity to infant distress, rather than sensitivity to nondistress, as the effects of parental attachment histories may be most evident when parents are challenged by attachment stressors requiring a parenting response (Thompson, 1997). Not only has parental sensitivity to distress been shown to predict infant attachment security more reliably than sensitivity to nondistress (McElwain & Booth-LaForce, 2006), it also has shown stronger associations with infants' own physiological self-regulation during distress (Conradt & Ablow, 2010).

The results linking women's physiology to maternal sensitivity were mixed and suggest that there may be differences in how aspects of women's autonomic nervous system (ANS) predict maternal sensitivity, as well as variation in the ways in which attachment and physiology operate together to predict behavior. Following established (Barron & Kenny, 1986) and more recent criteria (Kraemer et al., 2001), we also tested for mediated and moderated processes between our prenatal predictors.

Cardiac arousal in the form of increased HR was not associated with greater sensitivity, nor did prenatal HR interact with attachment status to predict behavior. Although prior research has documented HR elevations in response to infant cry (Frodi & Lamb, 1980), the level of cardiac arousal evoked by our cry paradigm was either too small or incommensurate with the level of cardiac response needed to support maternal sensitivity (Bornstein & Suess, 2000). Likewise, women's electrodermal reaction to infant cry did not predict later maternal response to their distressed infant. Instead, women's SCL response to cry appeared to moderate the association between attachment status and sensitivity such that women classified as insecure–dismissing who responded to the complex cry stimulus with increased SCL were significantly less sensitive with their distressed infant than the insecure–dismissing women who showed low SCL, or secure women at either level of SCL. As such, high SCL in response to infant cry may operate as added vulnerability, whereas low SCL may operate as a protective factor against less sensitive parenting in the context of insecure–dismissing attachment.

In contrast to the HR and SCL cry results, we found that women's prenatal RSA levels predicted individual differences in maternal sensitivity and lower RSA in response to both cry stimuli anticipated higher maternal sensitivity, whereas higher RSA predicted less sensitivity. These findings converge with Porges (2007) in suggesting that maternal vagal self-regulation during stressful parent–child interactions plays a significant role in organizing effective parenting behaviors (Hill-Soderlund et al., 2008; Mills-Koonce et al., 2007; Moore et al., 2009). We attempted to extend this line of work by testing the extent to which vagal regulation might account for the links between attachment and maternal sensitivity. Although our initial analyses showed that lowered RSA in response to infant cry was characteristic of secure–autonomous women, and that elevated RSA was characteristic of insecure–dismissing women, we did not find support for the hypothesis that women's prenatal RSA response mediated (or moderated) the links between attachment status and postnatal maternal sensitivity. As such, we cannot conclude that mother's vagal regulation in response to infant cries explains why mothers with secure attachment representations are more sensitive than insecure–dismissing women when responding to their distressed babies. The lack of mediation over the course of the first year does not rule out the possibility that women's concurrent vagal response to their own infant's distress is not mediating the effects of their attachment representations on how they respond to their own distressed infant. Nevertheless, these data suggest that both prenatal attachment status and RSA in response to an unfamiliar infant's cry signals contribute uniquely to mothers' sensitivity and responsiveness to their own infants' distress 1 year later.

These findings and their interpretations must be considered in the context of several limitations. First, given this sample comprised primarily White, middle- to upper-middle-class women, the generalizability of these findings is limited. Greater cultural and socioeconomic diversity should be sought before biobehavioral findings of the type reported here can be applied more broadly. The size of our sample, both overall and in terms of the number of women with secure and insecure representations of attachment, was another limitation. Not only did this yield modest cell sizes for our analyses, but we had too few cases to explore the biobehavioral implications of the AAI's two other primary classifications (insecure–preoccupied and unresolved with respect to trauma). Still, despite our modest sample size, the effect sizes reported herein were noteworthy. In particular, interactions involving attachment status and the different segments of our two cry stimuli consistently yielded meaningful effect sizes (Cohen, 1992). Provided these results can be replicated, these are not trivial levels of association and support the contention that women's physiological reactions to infant cry signals may differ partially as a function of their attachment history.

Conclusions

Variation in caregivers' responsiveness to infants' signals can lay the groundwork for subsequent emotion regulation abilities. Much of this early programming transpires within the context of the primary attachment relationship(s), where infants' bids for regulation may or may not be met with effective parenting. Consequently, caregivers' psychophysiological reactions to attachment cues are an important focus of research that could help to elucidate individual differences in sensitive parenting and, in turn, emotion development in children. Our goal was to contribute to an understanding of these potential processes both before women actually gave birth to their first child. We found that expectant women manifested reasonably distinct patterns of autonomic response to infant cry stimuli based on their AAI classification and that these attachment-linked differences in women's sympathetic response, in particular, predicted differences in maternal sensitivity once their child was born. These data suggest that primiparous women who were classified as secure–autonomous on the AAI showed an autonomic response that was consistent with approach-related behavior. In contrast, primiparous women classified as insecure–dismissing on the AAI showed an autonomic response that was more consistent with behavioral inhibition. Once mothers, insecure–dismissing women who responded highest on autonomic markers of behavioral inhibition were observed as the least sensitive with their own 9-month-old children. Although replication will be essential, these data appear to imply that exposure to an infant's cry is differentially arousing for women with secure versus insecure models of attachment. As these findings were derived with first-time expectant women during their final trimester of pregnancy, these data point to a potentially fruitful period of intervention. Universal psychoeducational programs that were embedded within standard prenatal health care practices might sensitize expectant parents to the importance of optimal response patterns as well as to their own characteristics that might interfere with optimal parenting.

References

Adam, E. K., Gunnar, M. R., & Tanaka, A. (2004). Adult attachment, parent emotion, and observed parenting behavior: Mediator and moderator models. *Child Development, 75*, 110–122.

Ainsworth, M. D. S., Blehar, M. C., Water, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.

Atkinson, L., Goldberg, S., Ravel, V., Pederson, D., Benoit, D., Moran, G., et al. (2005). On the relation between maternal state of mind and sensitivity in the prediction of infant attachment security. *Developmental Psychology, 41*, 42–53.

Atkinson, L., Niccols, A., Paglia, A., Coolbear, J., Parker, K. C. H., Poulton, L., et al. (2000). A meta-analysis of time between maternal sensitivity and attachment assessments: Implications for internal working models in infancy/toddlerhood. *Journal of Social and Personal Relationships, 17*, 791–810.

Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology, 37*, 13–20.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Beauchaine, T. P. (2001). Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and Psychopathology, 13*, 183–214.

Beauchaine, T. P., Gatzke-Kopp, L., & Mead, H. K. (2007). Polyvagal theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology, 74*, 174–184.

Beijersbergen, M. D., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Jeffer, F. (2008). Stress regulation in adolescents: Physiological reactivity during the Adult Attachment Interview and conflict interaction. *Child Development, 79*, 1707–1720.

Berntson, G. G., Norman, G. J., Hawkley, L. C., & Cacioppo, J. T. (2008). Cardiac autonomic balance versus cardiac regulatory capacity. *Physiology, 45*, 643–652.

Biringen, Z. (1990). Direct observation of maternal sensitivity and dyadic interactions in the home: Relations to maternal thinking. *Developmental Psychology*, *26*, 278–284.

Blokland, K., & Goldberg, S. (1988). *Attachment and expectant mothers' responses to infant emotion*. Paper presented at the biennial International Conference on Infant Studies, Washington, DC.

Bornstein, M. H., & Suess, P. E. (2000). Physiological self-regulation and information processing in infancy: Cardiac vagal tone and habituation. *Child Development*, *71*, 273–287.

Boulet, J., & Boss, M. (1991). Reliability and validity of the brief symptom inventory. *Journal of Consulting and Clinical Psychology*, *3*, 433–437.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss, sadness, and depression*. New York: Basic Books.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology*, *17*, 271–301.

Broth, M. R., Goodman, S. H., Hall, C., & Raynor, L. C. (2004). Depressed and well mothers' emotion interpretation accuracy and the quality of mother-infant interaction. *Infancy*, *6*, 37–55.

Bugental, D. B., Lewis, J. C., Lin, E., Lyon, J., & Kopeikin, H. (2000). In charge but not in control: The management of authority-based relationships by those with low perceived power. *Developmental Psychology*, *35*, 1367–1378.

Cacioppo, J. T., & Berntson, G. G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science*, *8*, 133–137.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Conradt, E., & Ablow, J. C. (2010). Infant physiological response to the still-face paradigm: Contributions of maternal sensitivity and infants' early regulatory behavior. *Infant Behavior and Development*, *33*, 251–265.

Crowe, H. P., & Zeskind, P. S. (1992). Psychophysiological and perceptual responses to infant cries varying in pitch: Comparison of adults with low and high scores on the Child Abuse Potential Inventory. *Child Abuse and Neglect*, *16*, 19–29.

Crowell, J. A., & Feldman, S. S. (1988). Mothers' internal models of relationships and children's behavioral and developmental status: A study of mother–child interaction. *Child Development*, *59*, 1273–1285.

Dawson, M. E., Schell, A. M., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 200–223). New York: Cambridge University Press.

Del Vecchio, T., Walter, A., & O'Leary, S. G. (2009). Affective and physiological factors predicting maternal response to infant crying. *Infant Behavior and Development*, *32*, 117–122.

Derogatis, L. R. (1975). *Brief Symptom Inventory*. Baltimore: Clinical Psychometric Research.

Dias, P., Soares, I., Klein, J., Cunha, J. P. S., & Roisman, G. I. (2011). Autonomic correlates of attachment insecurity in a sample of women with eating disorders. *Attachment & Human Development*, *13*, 155–167.

Dozier, M., & Kobak, R. R. (1992). Psychophysiology in attachment interviews: Converging evidence for deactivating strategies. *Child Development*, *63*, 1473–1480.

Fowles, D. C. (1988). Physiology and psychopathology: A motivational approach. *Psychophysiology*, *25*, 373–391.

Fox, N. A., & Hane, A. A. (2008). Studying the biology of human attachment. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 217–240). New York: Guilford Press.

Frodi, A. M., & Lamb, M. E. (1980). Infants at risk for child abuse. *Infant Mental Health Journal*, *1*, 240–247.

Galles, S. J., Miller, A., Cohn, J. F., & Fox, N. A. (2002). Estimating parasympathetic control of heart rate variability; two approaches to quantifying vagal tone. *Psychophysiology*, *39*, S37.

George, C., Kaplan, N., & Main, M. (1996). *Adult Attachment Interview*. Unpublished manuscript, University of California at Berkeley.

Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology and health. *Journal of Personality and Social Psychology*, *63*, 221–233.

Gray, J. A. (1975). *Elements of the two-process theory of learning*. New York: Academic Press.

Groh, A. M., & Roisman, G. I. (2009). Adults' autonomic and subjective emotional responses to infant vocalizations: The role of secure base script knowledge. *Developmental Psychology*, *45*, 889–893.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, *9*, 87–108.

Grossman, P., & Kollai, M. (1993). Respiratory sinus arrhythmia, cardiac vagal tone, and respiration: Within- and between-individuals relations. *Psychophysiology*, *30*, 486–495.

Haley, D. W., & Stansbury, K. (2003). Infant stress and parent responsiveness: Regulation of physiology and behavior during still-face and reunion. *Child Development*, *74*, 1534–1546.

Ham, J., & Tronick, E. Z. (2006). Infant resilience to the stress of the still-face. *Annals of the New York Academy of Sciences*, *1094*, 83–104.

Hesse, E. (2008). The adult attachment interview: Protocol, method of analysis, and empirical studies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 552–598). New York: Guilford Press.

Hill-Soderlund, A. L., Mills-Koonce, W. R., Propper, C., Calkins, S. D., Granger, D. A., Moore, G. A., et al. (2008). Parasympathetic and sympathetic responses to the strange situation in infants and mothers from avoidant and securely attached dyads. *Developmental Psychobiology*, *50*, 361–376.

Hofer, M. A. (2006). Psychobiological roots of early attachment. *Current Directions in Psychological Science, 15*, 84–88.

Holland, A. S., & Roisman, G. I. (2010). Adult attachment security and young adults' dating relationships over time: Self-reported, observational, and physiological evidence. *Developmental Psychology, 46*, 552–557.

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158*, 848–856.

Lin, E. K., Bugental, D. B., Turek, V., Martorell, G. A., & Olster, D. H. (2002). Children's vocal properties as mobilizers of stress-related physiological responses in adults. *Personality and Social Psychology Bulletin, 28*, 346–357.

Main, M. (2000). The organized categories of infant, child, and adult attachment: Flexible vs. inflexible attention under attachment-related stress. *Journal of the American Psychoanalytic Association, 48*, 1055–1096.

Main, M., & Goldwyn, R.. (1998). *Adult Attachment Interview scoring and classification manual* (6th version). Unpublished manuscript, University of California, Berkeley.

Marcos, H.. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language, 14*, 255–268.

McElwain, N. L., & Booth-LaForce, C. (2006). Maternal sensitivity to infant distress and nondistress as predictors of infant-mother attachment security. *Journal of Family Psychology, 20*, 247–255.

Mills-Koonce, W. R., Gariepy, J. L., Propper, C., Sutton, K., Calkins, S., Moore, G., et al. (2007). Infant and parent factors associated with early maternal sensitivity: A caregiver-attachment systems approach. *Infant Behavior and Development, 30*, 114–126.

Moore, G. A., & Calkins, S. D. (2004). Infants' vagal regulation in the still-face paradigm is related to dyadic coordination of mother-infant interaction. *Developmental Psychology, 40*, 1068–1080.

Moore, G. A., Hill-Soderlund, A. L., Propper, C. B., Calkins, S. D., Mills-Koonce, W. R., & Cox, M. J. (2009). Mother-infant vagal regulation in the face-to-face still-face paradigm is moderated by maternal sensitivity. *Child Development, 80*, 209–223.

Murray, L., Fiori-Cowley, A., Hooper, R., & Cooper, P. (1996). The impact of postnatal depression and associated diversity on early mother-infant interactions and later infant outcomes. *Child Development, 67*, 2512–2526.

Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage: A polyvagal theory. *Psychophysiology, 32*, 301–318.

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology, 74*, 116–143.

Raval, V., Goldberg, S., Atkinson, L., Benoit, D., Myhal, N., Poulton, L., et al. (2001). Maternal attachment, maternal responsiveness and infant attachment. *Infant Behavior & Development, 24*, 281–304.

Roisman, G. I. (2007). The psychophysiology of adult attachment relationships: Autonomic reactivity in marital and premarital interactions. *Developmental Psychology, 43*, 39–53.

Roisman, G. I., Tsai, J. L., & Chiang, K.-H. S. (2004). The emotional integration of childhood experience: Physiological, facial expressive, and self-reported emotional response during the Adult Attachment Interview. *Developmental Psychology, 40*, 776–789.

Rottenberg, J., Wilhelm, F. H., Gross, J. J., & Gotlib, I. H. (2002). Respiratory sinus arrhythmia as a predictor of outcome in major depressive disorder. *Journal of Affective Disorders, 71*, 265–272.

Spangler, G., & Grossmann, K. E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Development, 64*, 1439–1450.

Thompson, R. A. (1997). Sensitivity and security: New questions to ponder. *Child Development, 68*, 595–597.

Tronick, E. Z., & Cohn, J. F. (1989). Infant–mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child Development, 60*, 85–92.

van IJzendoorn, M. H. (1995). Adult attachment representations, parental responsiveness and infant attachment: A meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychological Bulletin, 117*, 387–403.

Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. *Psychological Bulletin, 109*, 147–151.

Ward, M. J., & Carlson, E. A. (1995). Associations among adult attachment representations, maternal sensitivity, and infant-mother attachment in a sample of adolescent mothers. *Child Development, 66*, 69–79.

Weisenfeld, A. R., Malatesta, C. A., Whitman, P. B., Granrose, C., & Vili, R. (1985). Psychophysiological response of breast- and bottle-feeding mothers to their infants' signals. *Psychophysiology, 22*, 79–86.

Wood, R. M., & Gustafson, G. W. (2001). Infant crying and adults' anticipated caregiving responses: Acoustic and contextual influences. *Child Development, 72*, 1287–1300.

Zeanah, C. H., Benoit, D., Barton, M., Regan, C., Hirshberg, L. M., & Lipsitt, L. (1993). Representations of attachment in mothers and their one-year-old infants. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 278–286.

Zeskind, P. S., & Lester, B. M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. *Child Development, 49*, 580–589.

Zeskind, P. S., & Lester, B. M. (2001). Analysis of infant crying. In L. T. Singer & P. S. Zeskind (Eds.), *Biobehavioral assessment of the infant* (pp. 149–166). New York: Guilford Press.

Zeskind, P. S., & Marshall, T. R. (1988). The relations between variations in pitch and maternal perceptions of infant crying. *Child Development, 59*, 193–196.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1392–1406

Visual Attention to Global and Local Stimulus Properties in 6-Month-Old Infants: Individual Differences and Event-Related Potentials

Maggie W. Guy, Greg D. Reynolds, and Dantong Zhang
University of Tennessee

Event-related potentials (ERPs) were utilized in an investigation of 21 six-month-olds' attention to and processing of global and local properties of hierarchical patterns. Overall, infants demonstrated an advantage for processing the overall configuration (i.e., global properties) of local features of hierarchical patterns; however, processing advantages were found to vary based on individual differences in look duration. Short-looking infants showed differences in the negative central ERP component and the late slow wave (LSW) indicating greater attention to and discrimination of changes in global properties. Analysis of the LSW revealed that long-looking infants discriminated changes in local features, but did not discriminate changes in global properties of visual stimuli. These findings indicate that short- and long-looking infants utilize different approaches when processing hierarchical patterns.

Global and local processing have been examined extensively in research with adult participants (for review, see Kimchi, 1992). Local features refer to the fine details or individual elements contained within the stimulus, whereas global properties refer to the overall configuration or pattern formed by the local features of the stimulus. Although there is a large body of work on configural and featural processing of faces (e.g., Maurer, Le Grand, & Mondloch, 2002), much of the research in the area has utilized hierarchical patterns to systematically examine processing of global and local stimulus properties. Hierarchical patterns are patterns in which larger figures are constructed from an arrangement of smaller figures. Navon-type figures (e.g., Navon, 1977) are commonly used examples that consist of large letters constructed from either a set of the same (congruent) smaller letter or a set of a different (incongruent) smaller letter. The larger patterns are considered higher level units than the smaller letters or features, and units that are higher on the hierarchy are considered more global than lower level units (Kimchi, 1992).

Under this framework, Navon (1977) proposed the global-precedence hypothesis, stating that perceptual processing proceeds from the global to the local level, progressively moving toward more and more fine-grained analysis. Support for global precedence comes from studies showing "global advantage" through faster discrimination of target stimuli based on global properties, as well as studies showing "global interference" when global information is incongruent with target stimuli (Kimchi, 1988; Miller, 1981; Navon, 1977, 1981). Recent adult research has found global precedence effects in the auditory modality as well (Bouvet, Rousset, Valdois, & Donnadieu, 2011; List, Justus, Robertson, & Bentin, 2007; Sanders & Poeppel, 2007). However, factors that have been found to influence the direction of precedence effects include relative size (Navon & Norman, 1983), local element density (Kimchi, 1983, 1990; Martin, 1979), and goodness of form (Hoffman, 1980). For example, although global precedence effects consistently occur when hierarchical patterns are composed of many local elements, a shift to local precedence effects occurs to patterns composed of few local elements (Kimchi, 1983, 1990; Martin, 1979). Findings indicate that this shift occurs at 7 ± 2 elements for both adults (Kimchi, 1983; Kimchi & Palmer, 1982) and children as young as 3 years of age (Kimchi, 1990).

The authors wish to thank Kelsie Bottiggi, Kristin Dye, Angela McLemore, and Lindsay Robinson for their assistance with participant recruitment and data collection. We also wish to thank the anonymous reviewers for their helpful comments and suggestions on a previous version of this manuscript. We are especially grateful to the parents and infants who participated in this study.

Correspondence concerning this article should be addressed to Greg D. Reynolds, Department of Psychology, University of Tennessee, Knoxville, TN 37996. Electronic mail may be sent to greynolds@utk.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0020
DOI: 10.1111/cdev.12053

Dukette and Stiles (2001) found developmental differences in processing hierarchical stimuli based on stimulus density (i.e., the distance between local elements). When stimulus density was reduced by eliminating some of the local elements while maintaining the original overall size of the stimulus, 4-year-olds showed a local advantage whereas 6-year-olds and adults maintained a global advantage. In a discrimination task using objects and nonobjects to form hierarchical patterns, Poirel, Mellet, Houdé, and Pineau (2008) found evidence of a progression from a local advantage at 4 years of age to adult-like global advantages at 9 years of age. Kimchi, Hadad, Behrmann, and Palmer (2005) investigated perceptual processing of many- and few-element patterns in participants ranging in age from 5 to 23 years. The authors reported mature processing of global configurations of many-element patterns and mature processing of local features of few-element patterns at 5 years of age. However, significant improvements were found from 5 to 10 years of age in processing global configurations of few-element patterns and in processing local features of many-element patterns.

Research with infants indicates that 3- to 4-month-old infants are capable of processing both global and local properties of hierarchical patterns (Ghim & Eimas, 1988), and global precedence effects are present at 3 months of age when processing many-element patterns (Frick, Colombo, & Allen, 2000). Taken together, these findings indicate that both global and local precedence effects are based on a dominant tendency under certain conditions to direct attention to and process one level of stimulus properties over the other level (Pomerantz, 1983) as opposed to being based on an obligatory sequence of perceptual processing (Navon, 1977).

Several studies have examined global and local processing in infancy in relation to individual differences in visual attention (Colombo & Mitchell, 1990; Colombo, Mitchell, Coldren, & Freeseman, 1991; Freeseman, Colombo, & Coldren, 1993; Frick & Colombo, 1996; Frick et al., 2000). Shorter look durations in infancy have been associated with: an increase in responsiveness to novelty, advanced motor development, and intelligence test performance up to 8 years of age (Bornstein & Sigman, 1986; Colombo, 1993; Colombo & Mitchell, 1990; Fagan, 1984; Rose, Slater, & Perry, 1986). Research in infant visual attention reveals stable individual differences in look duration (Colombo, Mitchell, O'Brien, & Horowitz, 1987) and findings indicate short-looking infants are more likely to demonstrate recognition memory for a previously viewed stimulus than long-looking infants (Colombo & Mitchell, 1990; Colombo, Mitchell, & Horowitz, 1988; Colombo et al., 1991; Freeseman et al., 1993).

Colombo (1995) proposed that short lookers might utilize a more efficient visual intake strategy than long lookers, resulting in faster and more thorough processing of stimulus properties. Specifically, short lookers may utilize a global processing strategy and long lookers may utilize a local processing strategy (Colombo et al., 1991). For example, Freeseman et al. (1993) tested 4-month-olds using the paired-comparison procedure with hierarchical patterns. The stimuli consisted of geometric shaped configurations (diamond or hourglass patterns) composed of 13 upper-case letters (*N* or *Z*). After 10 s of familiarization, short lookers showed novelty preferences based on discriminating changes in global properties. After 20 s of familiarization, they were also able to discriminate based on changes in local features. Long lookers did not show evidence of discrimination until acquiring 40 s of familiarization, at which point they were able to discriminate based on global properties, but not local features. The results of this study suggest that short and long lookers utilize the same global to local processing sequence in this task, yet short lookers process information more efficiently.

However, a follow-up study by Colombo, Freeseman, Coldren, and Frick (1995) yielded results that were inconsistent with this conclusion. Four-month-old infants were familiarized with the same stimuli utilized by Freeseman et al. (1993). After familiarization, infants were presented with a paired comparison using two novel stimuli. One of the stimuli contained familiar global properties and novel local features; the other stimulus contained a novel global pattern composed of the familiar local features. The results were consistent with previous findings (Freeseman et al., 1993) for short lookers but not long lookers. After 20 s of familiarization, short lookers showed a preference based on global properties. With 30 s of familiarization, short lookers shifted to demonstrating a preference based on local features. This was not the case for long lookers, who did not show a preference toward novel global or novel local properties after 20, 30, or 40 s of familiarization. A preference did not emerge until long lookers were exposed to 50 s of familiarization, at which point they showed a preference for the stimulus composed of novel local features. Long lookers did not show a preference based on global properties at any length of familiarization. Similarly, Colombo, Frick, Ryther, and Gifford (1996), found 4-month-old short lookers were able to recognize forms based on complementary contour, a process that requires global perception of the overall form and controls for potential discrimination of global forms based on perception of local features. In contrast, 4-month-old long lookers did not demonstrate recognition of forms based on complementary contour.

Taken together, these findings demonstrate local precedence effects for long lookers across a range of tasks and suggest that unlike short lookers, long lookers' selective attention is dominated at the local level in infancy. Findings from a study on visual-manual object examination in 8-month-olds (Macchi Cassia & Simion, 2002) provide support for this possibility. Short-looking infants looked longer at objects with novel-global properties and long-looking infants looked longer at objects with novel-local features. However, in a follow-up study (Frick et al., 2000) to Colombo et al. (1995), both short- and long-looking 3-month-olds showed evidence of processing global properties of visual patterns prior to local features. Thus, the visual intake strategy used by long-looking infants in particular remains unclear and may vary with age and task.

The event-related potential (ERP) is an excellent tool for investigating cognitive processing in the preverbal infant, which could provide insight into the processing strategies used by short and long lookers. ERPs represent voltage oscillations in the electroencephalogram (EEG) that are time-locked with the occurrence of a specific event such as stimulus onset (Fabiani, Gratton, & Coles, 2000; Picton et al., 2000). ERP components are identified in the averaged ERP waveform that are associated with subprocesses involved in perceptual or cognitive processing. Two ERP components that are particularly relevant for research on infant visual processing include the negative central (Nc) component, associated with infant visual attention, and the late slow wave (LSW), associated with stimulus processing and recognition memory (De Haan & Nelson, 1997; Reynolds, Courage, & Richards, 2010; Reynolds & Richards, 2005; Snyder, 2010).

Nc is a negatively polarized deflection in the waveform that typically occurs from 350–750 ms after stimulus onset. High amplitude Nc has been observed to salient and novel stimuli and has been proposed to reflect attentional engagement, novelty detection, or a general orienting response (Carver, Bauer, & Nelson, 2000; Courchesne, Ganz, & Norcia, 1981; De Haan & Nelson, 1997; Nikkel & Karrer, 1994; Reynolds & Richards, 2005; Reynolds et al., 2010; Richards, 2003). Nc has been shown to gradually increase in amplitude over the 1st year of life and then decrease significantly by 3 years of age (De Haan, 2007). Support for the proposal that Nc reflects attentional engagement comes from studies on 4.5- to 7.5-month-olds showing that Nc is greater in amplitude when heart rate measures are indicative of attention compared to periods when heart rate measures indicate inattention (Reynolds et al., 2010; Richards, 2003). Furthermore, Nc is greater in amplitude toward a mother's face compared to a dissimilar looking stranger's face (De Haan & Nelson, 1997), and Nc is greater in amplitude toward an infant's visually preferred stimulus regardless of novelty or familiarity (Reynolds et al., 2010). Thus, Nc amplitude most likely reflects the magnitude of the attentional response, and high amplitude Nc may be indicative of activation of a general arousal system involved in attention (Reynolds, Courage, & Richards, in press; Reynolds et al., 2010).

The LSW occurs from 1,000 to 2,000 ms after stimulus onset over temporal, parietal, frontal, and central electrodes and may be positive or negative in polarity (De Haan, 2007). Researchers have proposed that a positive LSW may indicate ongoing stimulus processing, whereas a negative LSW may indicate novelty detection (De Haan, 2007; De Haan & Nelson, 1997). However, this pattern of results has not been consistently supported and both negative (Nelson & Collins, 1991, 1992; Quinn, Westerlund, & Nelson, 2006; Reynolds & Richards, 2005; Richards, 2003) and positive (De Haan & Nelson, 1997; Snyder, 2010; Snyder, Webb, & Nelson, 2002; Wiebe et al., 2006) going LSWs have been seen in response to irregular or novel stimuli. A consistent finding across studies is that there is a change in the amplitude of the LSW with repeated stimulus exposure that is proposed to reflect stimulus processing and recognition (De Haan & Nelson, 1997; Reynolds, Guy, & Zhang, 2011; Richards, 2003; Snyder, 2010).

Because the polarity and directional change of the LSW can vary across electrode sites and tasks, specific predictions regarding the direction of LSW effects can be problematic (De Haan, 2007). Thus, analysis of the LSW is most informative as an index of stimulus processing or recognition memory when stimuli varying in amount of exposure or familiarity are used within the same procedure. This allows for comparisons of differential LSW responsiveness based on stimulus type (e.g., novel compared to familiar). For example, Snyder (2010) found that infants who show significant changes in the amplitude of the LSW across early to late trials in a habituation procedure are more likely to show behavioral evidence of recognition (i.e., novelty preferences) compared to infants who do not show changes in LSW amplitude across early to late trials.

Reynolds et al. (2011) utilized ERPs to investigate neural correlates of individual differences in visual attention in 6- and 7.5-month-old infants. Infants were first exposed to 20 s of accumulated looking to a female face (used to determine looker type) followed by 20 s of accumulated looking to a color photograph of an object that served as the familiar stimulus in subsequent ERP testing. During ERP testing, infants were shown repeated presentations of the familiar object and novel objects. No differences were found between groups or stimulus types on the Nc component associated with infant attention. However, short lookers showed evidence of discriminating novel from the familiar stimuli by demonstrating significant differences in LSW amplitude to familiar compared to novel stimuli at midline frontal and right temporal electrodes. No differences were found in LSW amplitude based on stimulus type for long lookers. Taken together, these Nc and LSW findings further suggest that differences in behavioral performance found across looker type groups may be based on differences in processing approach as opposed to differences in overall level of attentional engagement.

In the current study, we utilized ERPs to investigate visual attention to global and local stimulus properties for short- and long-looking 6-month-old infants. We tested 6-month-olds exclusively because of the lack of age differences between 6- and 7.5-month-olds in Reynolds et al. (2011), and for consistency with previous ERP research on infant attention and memory (e.g., Ackles & Cook, 2007; De Haan & Nelson, 1997; Reynolds & Richards, 2005; Reynolds et al., 2010; Snyder, 2010). We used many-element hierarchical stimuli similar to those used by Colombo et al. (1995) and hypothesized that short lookers would demonstrate differential ERP responding between familiar and novel-global stimuli, and no differences in ERP responding between familiar and novel-local stimuli. This would be consistent with a global precedence when processing many-element hierarchical stimuli for short-looking infants similar to adults and older children (Kimchi, 1990, 1992; Poirel et al., 2008). It was hypothesized that long lookers would either show similar ERP responding to all stimulus types due to insufficient processing, or significant differences between ERPs to familiar and novel-local stimuli but no differences between familiar and novel-global stimuli. The latter set of findings would be consistent with a local precedence for long-looking infants on this task.

Method

Participants

A sample of 21, 6-month-old infants (12 m, 9 f) was tested. All infants were tested within 2 weeks of their 26-week birthdates. Only infants born full-term (at least 38 weeks gestation) without complications and of normal birth weight were recruited. Participants were drawn from a predominantly Caucasian and middle-class population. An additional 78 infants were tested, but not included in the final sample due to fussiness ($N = 25$), insufficient number of artifact free trials ($N = 42$), and technical problems ($N = 11$). This attrition rate is high but not out of the normal range for infant ERP research due to issues related to high levels of movement artifact in the EEG (DeBoer, Scott, & Nelson, 2007).

Apparatus

Participants were positioned on their parent's lap in a sound-attenuated room. They were seated 55 cm away from a 27″ color LCD monitor (Dell 2707 WFP; Dell, Round Rock, TX). To limit distraction during testing, the participant seating area was surrounded by black curtains on all sides except for the side located behind the participant. Room lighting remained on throughout testing. A digital camcorder (Sony DCR-HC28; Sony, Tokyo, Japan) was located just below the monitor for judging infant visual fixations. Fixations were judged online using a video feed to a computer in the experiment control room, adjacent to the testing room. The video was also recorded through use of Netstation software produced by Electrical Geodesics Incorporated (EGI; Eugene, OR). The Netstation was used to record EEG data and to synchronize this data with the video. The experimental procedure was controlled on a PC using E-Prime 2.0 software (Psychology Software Tools Inc., Sharpsburg, PA). The E-Prime program sent experimental events to the Netstation and utilized a single-clock system to time-lock these experimental events with the EEG and video data.

Visual Stimuli

Female face. A digital color photograph of an adult female face was presented in a 20° square on the monitor. This was used to obtain peak look duration and to determine looker type.

Geometric patterns. Test stimuli consisted of 240 black-and-white Navon-type hierarchical patterns (Navon, 1977), which were composed of 24 white, uppercase letters arranged against a black background (see Figure 1). Three different stimulus types were used: familiar, novel-global, and novel-local. Novel-global stimuli differed from the familiar stimulus in overall configuration or global pattern, but were composed of the same local features (i.e., upper case letters) as the familiar pattern. Novel-local stimuli were composed of different local elements than the familiar stimulus, yet these elements were configured in the same global pattern as the familiar stimulus. The 240 stimuli consisted of 15 different patterns. There were 16 exemplars of all 15 patterns, each composed of a different uppercase letter for the local features. Individual infants were only exposed to 30 of these stimuli during testing: 1 familiar, 14 novel-global, and 15 novel-local. The patterns were presented in a 20° square centered on the computer monitor.

Sesame Street characters. Videos of Sesame Street characters were used to regain infants' interest and fixation on the monitor after becoming distracted. The Sesame Street video covered a 15° square area centered on the monitor.

Procedure

Following the informed consent process, infants were held on a parent's lap approximately 55 cm from the center of the computer monitor. They were fitted with an EGI sensor net and impedances were measured. The experiment was then completed in three phases. The first phase was designed for consistency with previous research examining looker type (e.g., Colombo et al., 1995; Colombo & Mitchell, 1990; Colombo, Richman, Shaddy, Greenhoot, & Maikranz, 2001; Frick, Colombo, & Saxon, 1999; Reynolds et al., 2011). A color photograph of a female face was displayed on the monitor until the infant reached 20 s of accumulated looking toward the face. Participants' peak look during this phase was used to determine looker type (short or long looker). Previous studies have shown high within-session concordance between peak (i.e., longest) looks to a female face and peak looks during more extended viewing of a variety of stimulus types in infancy (Colombo et al., 1991; Reynolds, Zhang, & Guy, in press). Looking data were analyzed offline in order to determine peak look. The looker type groups were then determined by using a median split. This resulted in 11 short lookers and 10 long lookers.

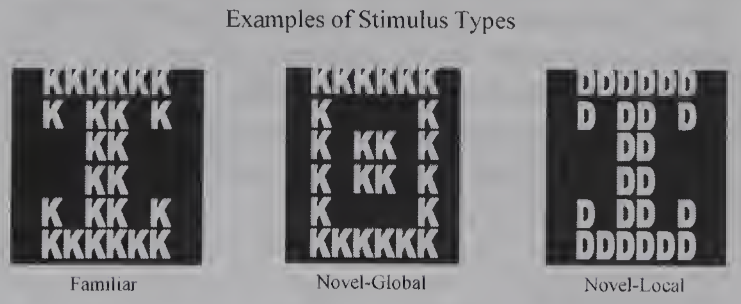


Figure 1. Examples of the three stimulus types used in the study. All stimuli were black-and-white geometric patterns (global patterns) composed of 24 uppercase letters (local features).

Following the presentation of the female face, the second phase began. This phase served as a familiarization phase. One of the geometric patterns was displayed on the monitor until the infant attained 20 s of accumulated looking. This stimulus was designated as the familiar stimulus for the remainder of the experiment. The pattern used for the familiar stimulus was randomly determined and varied between participants. The third phase consisted of the ERP test trials, during which the familiar, 15 novel-global, and 14 novel-local stimuli were briefly presented to each infant. For the novel-global presentations, stimuli were randomly selected with replacement from the 15 different patterns that shared the same local features (i.e., were composed of the same letter) as the familiar stimulus. For the novel-local presentations, stimuli were randomly selected with replacement from the 14 exemplars of the familiar pattern, which were composed of different letters (see Figure 1). The stimuli were presented for 500 ms, followed by a blank, black screen for a random duration of 2,000–2,500 ms. The minimum length of the interstimulus interval allowed for the analysis of the LSW that occurs from 1 to 2 s following stimulus onset and for a prestimulus interval prior to the subsequent stimulus onset. Novel-global, novel-local, and familiar stimulus presentations were equally distributed across trials. The stimuli were presented in random order in blocks of 30 stimulus presentations. Stimulus presentations were initiated only when the infant was judged to be looking at the monitor. During periods of distraction, the Sesame Street videos were presented as an attractor stimulus. In cases in which the attractor stimulus was used, there was always a blank screen for a minimum of 500 ms prior to the subsequent stimulus onset. Stimulus presentations continued for as long as the infant was not tired or fussy. Infants typically completed around a 100 trials, and testing lasted approximately 10 min on average.

Interobserver Reliability

Infant visual fixations to the female face were judged offline by two trained observers to determine looker type groups. Measures of interobserver reliability indicate high agreement between these observers (κ = .88). Infant fixations during the ERP testing phase were also judged offline. Trials in which the infants were judged to be looking away at any point during the stimulus presentation were not included in data analysis.

EEG Recording and Analyses

The EGI Geodesic EEG System 300 (GES 300) 128-channel system was used. The system consisted of the HydroCel Geodesic Sensor net, the NetAmps hardware, and the Netstation recording program. The 128-channel net used consisted of 124 electrodes mounted in a geodesic configuration of pedestals held in place with elastic connections. There were an additional 4 channels available for recording EOG or heart rate, which were not used in this experiment. Electrolytic sponges were located within the pedestals and the entire net was soaked in an electrolytic (saline-based) solution prior to use. Pedestals corresponding to the vertex, mastoids, and nasion locations were marked on the net and used to position the sensor net in relation to these anatomical landmarks while the elasticity of the net connections served to maintain the correct position of the pedestals corresponding to the remaining 120 electrodes. The average interelectrode distance of the scalp electrodes was 21 mm.

The proper placement of the sensor net results in electrode impedances of about 10–50 kΩ. The EGI system utilized high-impedance amplifiers containing 128 channels connected to a computer A/D card in a PowerPC-based computer system. The Netstation program was used for the A/D sampling, storing the data, the zero and gain calibration for each channel, and measuring impedances. The Netstation program also received serial communication from a Dell Workstation used to control the experimental protocol with the use of E-Prime 2.0 software. Communication between the two computers was temporally synchronized based on the sending of experimental information (e.g., trial type, trial onsets) from the experimental computer to the Netstation program using the E-Prime single-clock system. The sampling rate of the EEG was 250 Hz with 20K amplification. Band-pass filters were set from 0.3 to 100 Hz during recording. EEG recordings were referenced to the vertex. Following testing, EEG recordings were run through a 30-Hz low-pass filter and algebraically re-referenced to the average reference. The application and adjustment of the net took about 5–10 min, during this time a second experimenter distracted the infant with rattles and infant-directed speech to keep the infant in a positive state prior to testing.

The EEG recordings were inspected for artifacts (i.e., blinks, saccades, movement artifact, and drift) and poor recordings using the Netstation review system. Individual channels were marked bad within trials if these occurred. Segments in which more than 10% of the channels were marked bad were eliminated from the analysis. For trials that were retained for the ERP analysis, individual channels marked bad were replaced using a spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989; Srinivasan, Tucker, & Murias, 1998). Only those participants who contributed enough ERP trials per condition (i.e., at least 10 trials; De Boer et al., 2007) for stable ERP averages following EEG editing were included in the final data set. On average, infants contributed 15.14 trials (range = 10–29) in the familiar condition, 15.23 trials (range = 10–29) in the novel-global condition, and 15.77 trials (range = 10–27) in the novel-local condition.

ERP averages were calculated from 100 ms before stimulus onset through 2 s after onset. The segment length was chosen to allow for analysis of the Nc component and the LSW. The Nc component is typically located at midline frontal and central electrodes (i.e., Fz, Cz). The LSW has been observed at frontal, parietal, and temporal electrodes. We analyzed the mean data from clusters of electrodes of the EGI sensor net that correspond to these regions. As is standard practice in the field, the specific electrodes used in each cluster were determined based on past research and visual inspection of the grand average ERP waveforms (DeBoer et al., 2007). Nc peak (i.e., minimum) amplitude was analyzed from 350 to 750 ms following stimulus onset at midline frontal (5, 6, 11, and 12), and central electrode clusters (7, 31, 80, and 106). For the LSW, mean amplitude from 1 to 2 s following stimulus onset was analyzed separately at frontal (midline: 4, 10, 11, 16, 18, and 19; left: 24, 27, 28, 33, and 34; right: 116, 117, 122, 123, and 124), central (7, 31, 55, 80, and 106), parietal (midline: 61, 62, 67, 72, 77, and 78; left: 85, 86, and 92; right: 52, 53, and 60), and temporal (left posterior:

51, 58, 59, 64, and 65; right posterior: 90, 91, 95, 96, and 97) electrode clusters.

Design for Statistical Analysis

A priori planned comparisons were carried out using mixed analyses of variance (ANOVAs) with looker type (2: short looker, long looker) as a between-subjects factor, and stimulus type (3: familiar, novel-global, novel-local) as a within-subjects factor. We also analyzed electrode location as a within-subjects factor; the level for this factor varied by component and region. The Greenhouse-Geisser correction was used in cases of violations of the assumption of sphericity. For significant effects, follow-up analyses were done using one-way ANOVAs or paired-samples t tests (two-tailed). Effect sizes (η_p^2) are reported on all significant effects, and all significant tests are reported at $p < .05$.

Results

Looker Type Analysis

The twenty-one 6-month-old infants were split into short and long lookers using a median split on peak look to the photograph of the female face ($Mdn = 9.24$). This yielded 11 short-looking infants (peak look: $M = 7.53$ s) and 10 long-looking infants (peak look: $M = 14.65$ s). These look lengths are consistent with previous studies using accumulated looking to face stimuli with 6-month-old infants (e.g., Courage, Reynolds, & Richards, 2006; Reynolds et al., 2011).

ERP Grand Averages

The grand averages are shown by stimulus type across electrodes of interest in Figure 2. The Nc can be seen as a negative deflection at midline frontal and central electrodes occurring from 350 to 750 ms after stimulus onset. The LSW is most apparent at central and midline parietal electrodes, but can also be seen at temporal electrodes, occurring from 1 to 2 s after stimulus onset.

The Nc Component

Nc peak (minimum) amplitude was analyzed using a mixed ANOVA including midline electrodes (2: frontal, central) and stimulus type (3: familiar, novel-global, and novel-local) as within-subjects factors and looker type (2: short and long) as a between-subjects factor. There was a main

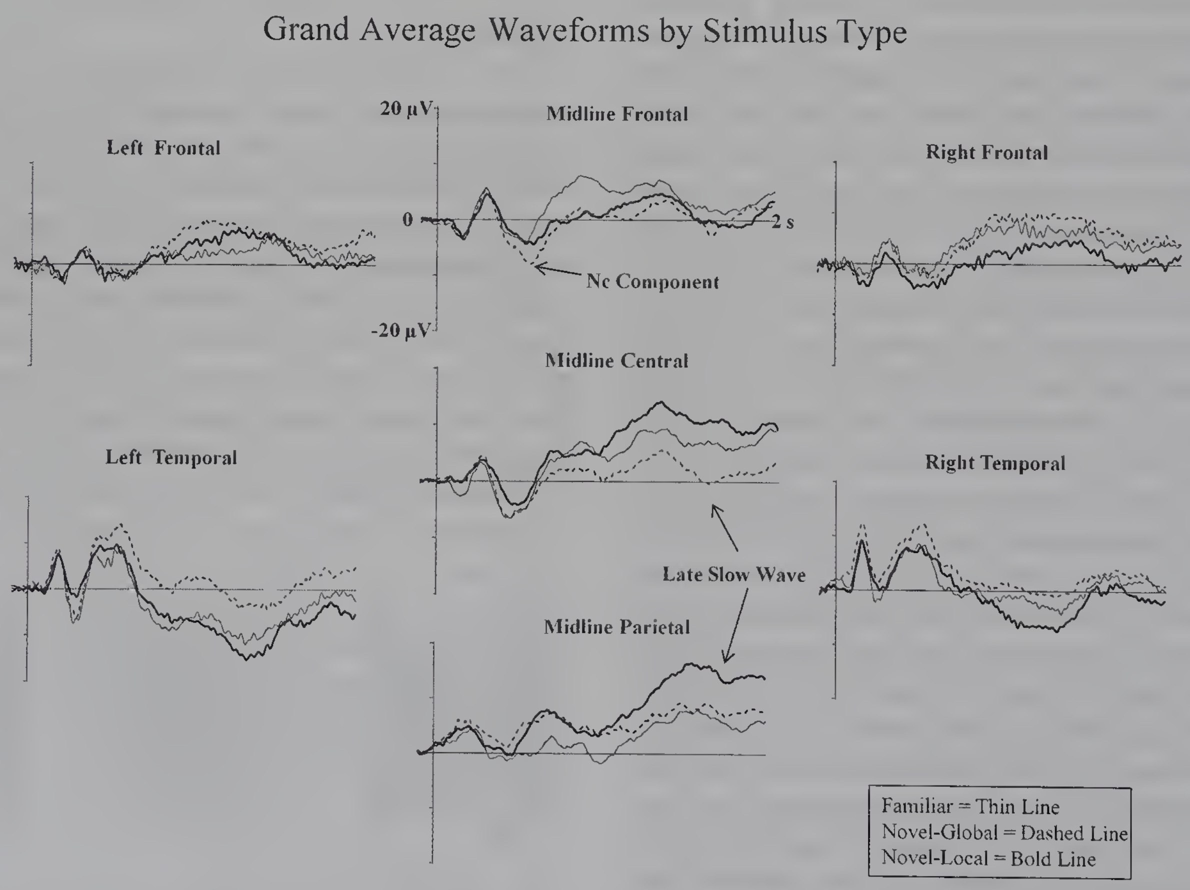


Figure 2. The grand average event-related potential waveforms by stimulus type and electrode. Waveforms are presented from representative leads for electrode clusters analyzed for experimental effects. The Nc (negative central) and late slow waves are indicated.

effect of stimulus type, $F(2, 38) = 5.80$, $p = .006$, $\eta_p^2 = .234$. As can be seen in Figure 2, infants demonstrated significantly greater amplitude Nc to novel-global ($M = -10.88$ μV, $SD = 7.68$ μV) compared to familiar ($M = -6.31$ μV, $SD = 9.09$ μV; $p = .002$) and novel-local ($M = -6.46$ μV, $SD = 10.05$ μV; $p = .011$) stimuli. No differences were found in Nc amplitude to familiar and novel-local stimuli ($p = .897$).

There was a marginal interaction of stimulus type and looker type, $F(2, 38) = 2.89$, $p = .068$, $\eta_p^2 = .132$. A priori planned comparisons were done separately for short and long lookers. For short lookers, there was a significant effect of stimulus type on Nc amplitude, $F(2, 20) = 8.715$, $p = .002$, $\eta_p^2 = .466$. As can be seen in Figure 3, short lookers demonstrated greater amplitude Nc to novel-global stimuli ($M = -11.66$ μV, $SD = 8.95$ μV) compared to familiar ($M = -4.91$ μV, $SD = 11.67$ μV; $p = .013$) and novel-local ($M = -3.93$ μV, $SD = 10.73$ μV; $p = .001$) stimuli. Their response to novel-local and familiar stimuli did not differ ($p = .631$). For long lookers, there were no differences in Nc amplitude based on stimulus type, $F(2, 18) = 0.521$, $p = .602$; $\eta_p^2 = .055$ (see Figure 3).

The LSW

LSW effects were examined using mixed ANOVAs on the mean amplitude from 1 to 2 s after stimulus onset separately for frontal, parietal, and temporal electrodes. At posterior temporal electrodes, a 3 (stimulus type) × 2 (electrode: left, right) × 2 (looker type) ANOVA revealed a significant main effect of stimulus type, $F(2, 38) = 3.608$, $p = .037$, $\eta_p^2 = .160$. The amplitude of response to novel-global stimuli ($M = -0.52$ μV, $SD = 7.52$ μV) was significantly different ($p = .007$) from novel-local stimuli ($M = -5.97$ μV, $SD = 7.98$ μV; see left and right temporal electrodes in Figure 2). LSW amplitude response to familiar stimuli ($M = -3.04$ μV, $SD = 9.64$ μV) was not significantly different from novel-global ($p = .193$) or novel-local ($p = .225$) stimuli.

At frontal electrodes, a 3 (stimulus type) × 3 (electrode: left, midline, right) × 2 (looker type) ANOVA revealed a three-way interaction of stimulus type, electrode, and looker type, $F(4, 76) = 2.822$, $p = .031$, $\eta_p^2 = .129$. Follow-up ANOVAs revealed a significant Stimulus Type × Looker Type interaction at midline frontal electrodes, $F(2, 38) = 3.977$, $p = .027$, $\eta_p^2 = .173$ (see Figure 3). Short lookers showed

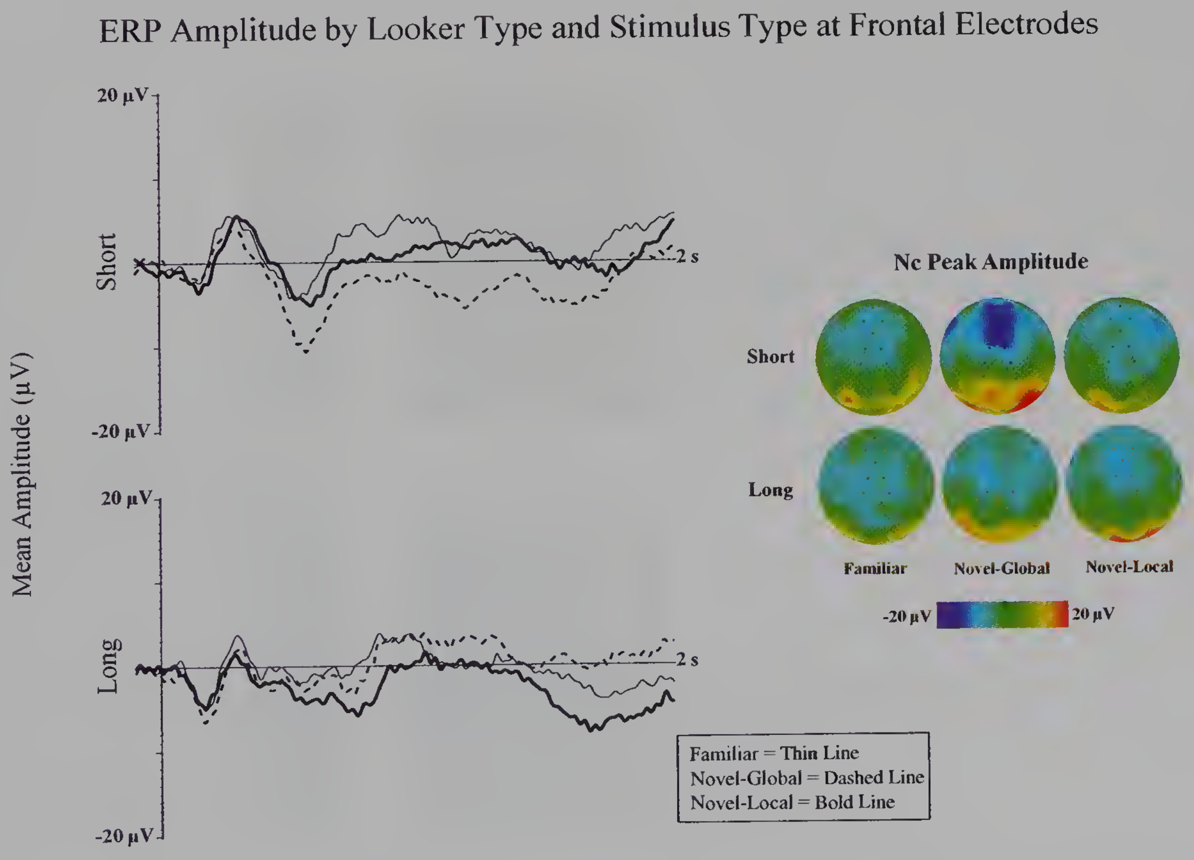


Figure 3. Event-related potential (ERP) amplitude is presented by stimulus type at frontal electrodes. The left panel shows ERP waveforms for short and long lookers at Fz. The shaded areas on the waveform plots indicate the time-windows for the Nc (350–750 ms) and LSW (1–2 s) analyses. The right panel shows topographical plots at peak amplitude Nc by stimulus type and looker type. The black dots in the topographical plots indicate the location of the electrodes used in the midline frontal and central electrode clusters.

significant differences ($p = .047$) to novel-global ($M = -2.04$ μV, $SD = 10.97$ μV) compared to familiar ($M = 4.67$ μV, $SD = 7.10$ μV) stimuli. There were no significant differences observed between novel-local ($M = 2.40$ μV, $SD = 10.77$ μV) and novel-global ($p = .168$) stimuli or novel-local and familiar ($p = .508$) stimuli for short lookers. Long lookers showed marginally significant differences ($p = .054$) between novel-local ($M = -3.91$ μV, $SD = 9.48$ μV) compared to novel-global ($M = 2.14$ μV, $SD = 7.10$ μV) stimuli. No differences were found between novel and familiar stimuli ($M = 0.20$ μV, $SD = 7.44$ μV).

There was also a significant effect of stimulus type at central electrodes, $F(1.756, 33.365) = 6.528$, $p = .005$, $\eta_p^2 = .256$. Infants demonstrated significant differences in LSW mean amplitude to novel-global ($M = 2.65$ μV, $SD = 6.17$ μV) compared to familiar ($M = 7.18$ μV, $SD = 6.11$ μV) and novel-local ($M = 9.29$ μV, $SD = 7.21$ μV) stimulus types ($ps = .01$ and $.004$ respectively; see central electrodes in Figure 2). There was a marginally significant interaction between looker type and stimulus type, $F(1.756, 33.365) = 2.981$, $p = .071$, $\eta_p^2 = .136$. A priori planned comparisons by looker type groups revealed a significant effect of stimulus type for short lookers, $F(2, 20) = 5.887$, $p = .010$, $\eta_p^2 = .371$ (see Figure 4).

Short lookers demonstrated significant differences in LSW amplitude between the novel-global condition ($M = 1.41$ μV, $SD = 6.70$ μV) compared to the familiar ($M = 9.92$ μV, $SD = 4.17$ μV; $p = .001$) and novel-local ($M = 9.39$ μV, $SD = 9.04$; $p = .025$ μV) conditions. There was a marginally significant stimulus type effect in the opposite direction for long lookers at central electrodes, $F(2, 18) = 3.258$, $p = .062$, $\eta_p^2 = .266$. This effect reached significance at parietal electrodes, $F(2, 18) = 4.599$, $p = .024$, $\eta_p^2 = .338$ (see Figure 5). Long lookers demonstrated significant differences in response to familiar ($M = 3.04$ μV, $SD = 6.46$ μV) compared to novel-local ($M = 11.15$ μV, $SD = 7.08$ μV) stimuli ($p = .004$). No other effects were significant at parietal electrodes.

Discussion

The goal of the current study was to investigate processing strategies in 6-month-old short and long lookers through the use of ERPs. We hypothesized that short lookers would display evidence of utilizing

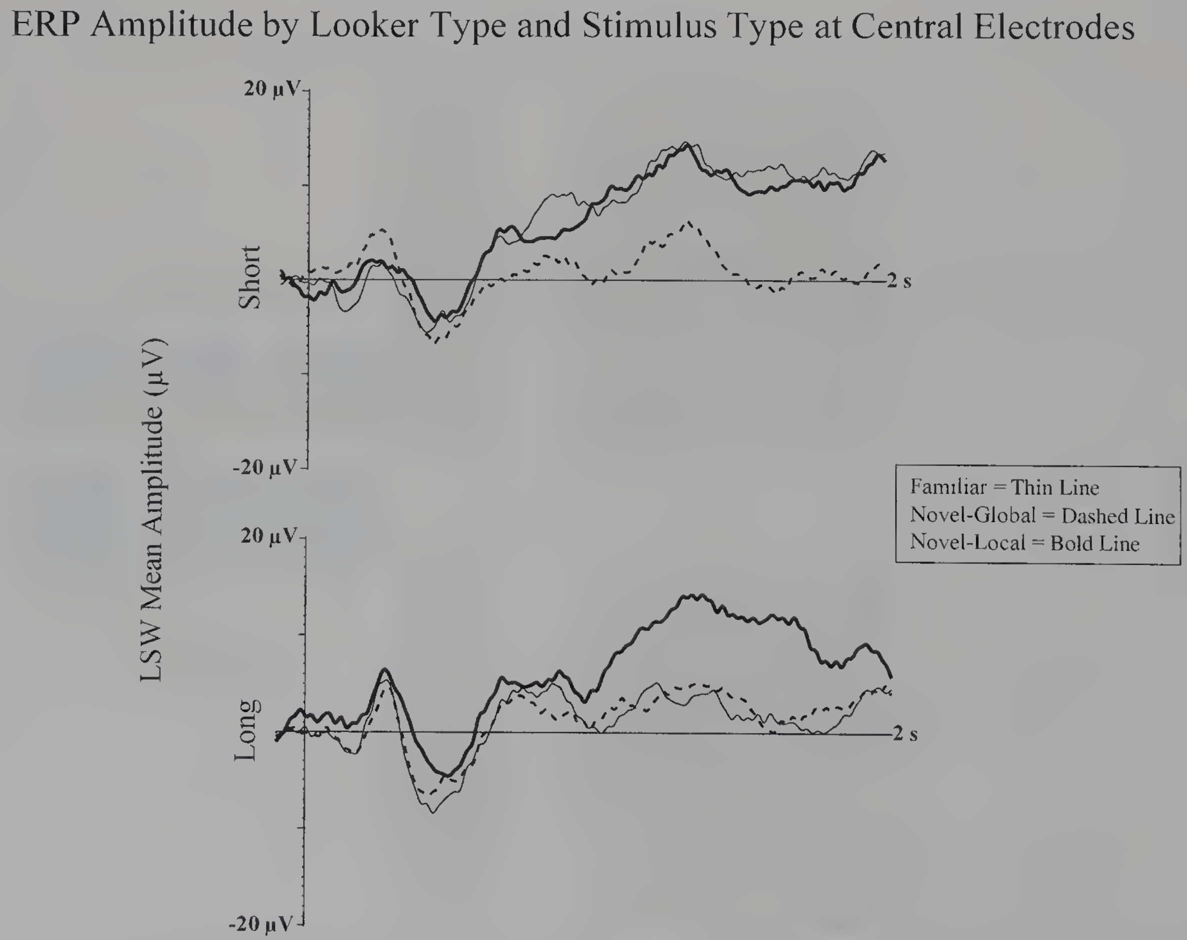


Figure 4. Event-related potential amplitude is presented by stimulus type and looker type at midline central electrode 80. The shaded areas on the waveform plots indicate the time windows for the Nc and LSW analyses.

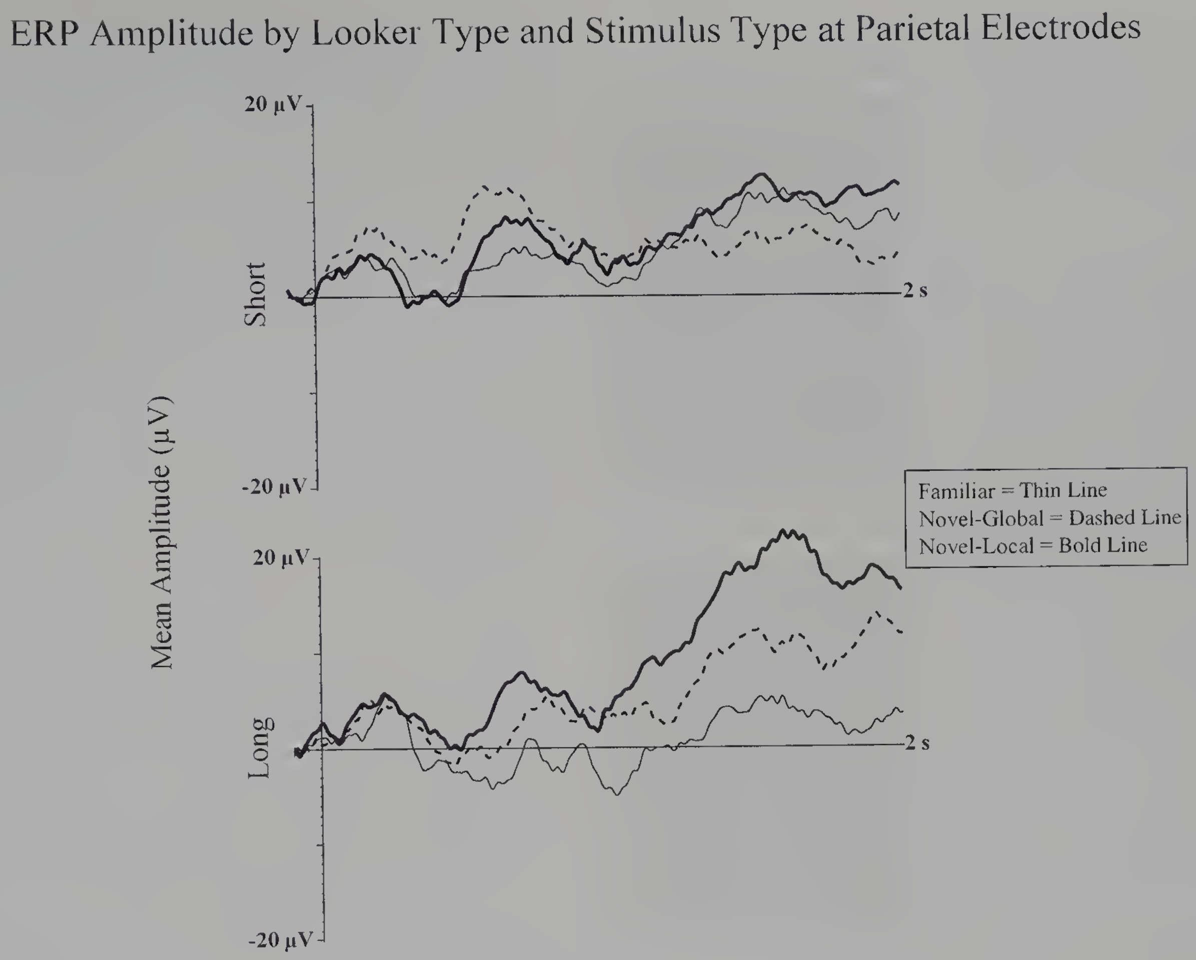


Figure 5. Event-related potential amplitude is presented by stimulus type and looker type at midline parietal electrode 61. The shaded area on the waveform plot indicates the time window for the LSW analysis.

a more mature, global precedence when processing hierarchical patterns, as evidenced by differential ERP responding to novel-global stimuli in comparison to familiar stimuli. Based on behavioral findings indicating that long lookers show novelty preferences based on local features after extended familiarization (Colombo et al., 1995), we predicted that any observed differences in ERP amplitude for long lookers would be between the familiar stimulus and novel-local stimuli. The results of our analyses of Nc and the LSW were consistent with these hypotheses and indicate that short and long lookers utilize different visual processing strategies in this task. Similar to older children and adults (Kimchi, 1983, 1990; Kimchi & Palmer, 1982), with 20 s of familiarization short lookers appear to demonstrate global precedence effects when processing many-element hierarchical patterns. In contrast, our findings indicate that long lookers demonstrate local precedence effects when processing these patterns after 20 s of familiarization.

The Nc Component

Our analysis of the Nc component, which is associated with infant attention (Ackles & Cook, 2007; Courchesne et al., 1981; De Haan, 2007; De Haan & Nelson, 1997; Reynolds & Richards, 2005) and visual preferences (Reynolds et al., 2010), revealed a main effect for stimulus type across midline frontal and central electrodes. Infants demonstrated greater amplitude Nc to novel-global stimuli compared to familiar stimuli. No differences were found between novel-local and familiar stimuli. Thus, consistent with previous behavioral work with 3- to 4-month-old infants (Frick et al., 2000; Ghim & Eimas, 1988), 6-month-old infants demonstrated global-precedence effects when processing these hierarchical patterns composed of many elements. A priori planned comparisons indicated short lookers accounted for this effect. Short lookers demonstrated significantly greater amplitude Nc to novel-global stimuli when compared to either familiar or novel-local stimuli. In contrast, no significant differences in Nc amplitude were found across stimulus types for long lookers. These differences were strongest at midline frontal electrodes (refer to Figure 3). Source localization studies indicate that frontal areas that are associated with attention processes in adults (e.g., anterior cingulate, inferior prefrontal cortex) are likely cortical sources of the Nc component (Reynolds & Richards, 2005;

Reynolds et al., 2010). These areas are involved in visual target detection and selective attention, and share reciprocal connections with areas involved in memory (Casey et al., 1997; Cohen, 1993; Nelson & Dukette, 1998; Rugg, Fletcher, Chua, & Dolan, 1999).

The current results indicate that with 20 s of familiarization, 6-month-old short lookers attend to changes in the global properties of visual patterns and do not attend to changes in the local features that comprise visual patterns. In contrast, with the same amount of familiarization, 6-month-old long lookers showed no evidence of attending to changes in global or local stimulus properties at the level of Nc, which suggests insufficient encoding for recognition of the familiar stimulus for this group. The results of the Nc analysis are consistent with behavioral findings indicating that short lookers process visual patterns more efficiently than long lookers, and short lookers demonstrate a global advantage similar to adults when processing hierarchical stimuli (Colombo et al., 1991; Colombo et al., 1995; Freeseman et al., 1993; Frick et al., 2000; Kimchi, 1992). However, because no differences were found on Nc amplitude across stimulus types for long lookers, the results of the Nc analysis do not provide insight into the visual processing strategy used by long-looking infants. In the adult literature, a temporal processing precedence has been found for global stimulus properties (Han, Yund, & Woods, 2003), with longer peak latencies for the N2 and P3 ERP components to local targets than to global targets (Han, Fan, Chen, & Zhuo, 1997). Some have speculated that Nc may be a precursor to the adult N2 component (Karrer & Monti, 1995; Nelson & Dukette, 1998).

Reynolds et al. (2011) found significant looker type effects on the LSW but no differences were found between short and long lookers on the Nc component. The lack of a main effect for looker type on Nc amplitude was replicated in the current study and suggests that short and long lookers do not differ in overall magnitude or depth of attention. Yet, the current finding that short lookers demonstrate differences in Nc amplitude based on stimulus type and long lookers do not suggests differences in the distribution of selective attention across these groups. Colombo et al. (2001) examined the relation between individual differences in look duration and heart rate measures of attention in infancy, and found that long-looking infants spend more time in sustained attention, the heart rate phase associated with active attention and information processing (Richards & Casey, 1991), than short-looking infants. However, this greater time spent in sustained attention did not equate to better recognition memory in subsequent behavioral testing. Because short lookers spent less time in sustained attention and still outperformed long lookers on a recognition task, the authors speculated that a measure that taps into the quality of attention, as opposed to temporal quantity, might reveal differences in attention across looker types. We believe that the pattern of combined effects of stimulus type on both Nc and the LSW for short lookers in the current study provides evidence of effective attention resulting in successful processing of global information.

The LSW

The LSW analyses provided evidence consistent with a global advantage in hierarchical processing for short lookers. At central electrodes (see Figure 4), short lookers showed differential LSW responding to the novel-global compared to either the familiar or novel-local stimuli. A significant difference in LSW amplitude between novel-global and familiar stimuli was also found at midline frontal electrodes for short lookers. No differences in LSW amplitude were found between familiar and novel-local stimuli for short lookers. Long lookers demonstrated LSW effects in the opposite direction. At central and midline frontal electrodes, there were marginally significant levels of discrimination between novel-local stimuli and the other stimulus types for long lookers. However, at parietal sites, long lookers demonstrated significant differences in LSW amplitude between the novel-local in comparison to either familiar or novel-global stimuli (see Figure 5). No significant differences were found between familiar and novel-global stimuli for long lookers.

The results of the current study are consistent with the hypothesis that short and long lookers utilize different visual processing strategies (Colombo et al., 1991). These data indicate that at 6 months of age, short lookers demonstrate a global advantage and long lookers demonstrate a local advantage when processing many-element hierarchical patterns with 20 s of familiarization. Taken together with findings from the behavioral literature (e.g., Colombo et al., 1991; Colombo et al., 1995; Freeseman et al., 1993; Frick et al., 2000), it appears that short-looking infants demonstrate global precedence, focusing attention primarily on global properties over local features of stimuli, by 3–6 months of age. These findings complement previous work demonstrating that children demonstrate adult-like global-precedence effects when processing many-element hierarchical patterns as early as 3 years of age (Kimchi, 1990; Kimchi et al., 2005; Poirel et al., 2008).

In a previous study (Freeseman et al., 1993), short lookers were able to discriminate both global and local changes in stimuli with 20 s of familiarization. We found no differential responsiveness to novel-local and familiar stimuli for short lookers in the current study. Procedural differences may account for the inconsistency in findings. For example, the stimuli used in the current study were composed of 24 letters and those used by Freeseman et al. (1993) were composed of 13 letters. Additionally, the repeated presentations of three stimulus types for several minutes following familiarization (required for the ERP analysis) may have interfered with short lookers' retention of local information in the current study. Given these inconsistencies, it should be noted that Freeseman et al. found that short lookers discriminated global changes with only 10 s of familiarization, thus both studies demonstrate global-precedence effects overall for short lookers.

Past studies have yielded inconsistent results regarding the processing strategy used by long-looking infants. Some findings indicate that long lookers utilize a global-to-local processing approach across a range of age groups in infancy (Freeseman et al., 1993; Frick et al., 2000), while other findings indicate that long lookers process in a local-to-global sequence across a broad age range in infancy (Colombo et al., 1995; Macchi Cassia & Simion, 2002). The current LSW findings further indicate that, unlike short-looking infants and adults, long-looking infants show local precedence effects when processing many-element hierarchical patterns. It is important to note that without manipulation of familiarization time, we cannot make conclusions regarding the temporal order of global and local processing for short and long lookers. However, our findings are consistent with global-precedence effects for short lookers and local-precedence effects for long lookers. We interpret these precedence effects as reflecting the dominance of either the global or local level at capturing selective attention under certain conditions (Pomerantz, 1983) as opposed to indicating an obligatory sequence of perceptual processing (Navon, 1977). Given this interpretation, it appears that short lookers utilize a mature processing approach and long lookers do not on this task.

The inconsistency in findings related to long lookers' approach to visual processing may be related to procedural differences and differences in age groups used across studies. For example, most of the studies supporting a global precedence for long lookers used 3- or 4-month-old infants (Colombo et al., 1991; Freeseman et al., 1993; Frick et al., 2000), whereas the studies with findings supporting a local precedence for long lookers used 4-month-olds (Colombo et al., 1995), 6-month-olds (current study), and 9-month-olds (Macchi Cassia & Simion, 2002). The latter studies also used a wider variety of stimuli and varied in procedure and level of analysis. We tested 6-month-olds in the current study as a direct follow-up to Reynolds et al. (2011), which tested 6- and 7.5-month-olds and found no age differences.

In the current study, long lookers showed the ability to recognize novel local stimulus characteristics after 20 s of familiarization, whereas in previous research conducted by Colombo and colleagues (Colombo et al., 1995; Freeseman et al., 1993) long lookers required 40 or 50 s to show any discrimination. We believe that this reflects a combination of an increase in processing efficiency from 4 to 6 months of age, and greater sensitivity of ERPs compared to looking measures in this context. Several studies integrating ERP and looking measures in infancy have found significant ERP effects and failed to find differences in looking time measures (see Reynolds & Guy, 2012, for a review). Given the characteristic looking style of long lookers, ERPs may provide a more sensitive measure of their discriminatory capabilities than preferential looking tasks.

Significant LSW effects were found to occur at different electrode sites for short (midline frontal, central) and long lookers (parietal). Snyder (2010) has speculated that the LSW reflects activity of medial temporal lobe structures associated with recognition memory. However, conclusions regarding cortical sources of ERP components cannot be made based on traditional EEG/ERP analyses, and unfortunately, practical and ethical concerns preclude the use of fMRI for basic research with infant participants (Reynolds & Richards, 2009). Furthermore, Colombo (1995) hypothesized that differences in visual scanning behavior that exist between short and long lookers may be based on delayed development of the posterior orienting system for long lookers compared to short lookers. The posterior orienting system, composed of parietal areas, the pulvinar nucleus of the thalamus, and frontal eye-fields, is involved in the voluntary control of eye movements (Posner & Petersen, 1990).

According to this hypothesis (Colombo, 1995), long lookers are delayed in their ability to voluntarily disengage and shift visual attention, thus demonstrating prolonged looks and a narrow focus of attention. Frick et al. (1999) found support for this hypothesis in a study that examined the latency to shift from a central to peripheral stimulus by short and long lookers. Long lookers were slower to shift fixation than short lookers, but only on trials where the central target remained, creating visual competition, and only on the trials when they were slowest to disengage. Combining looking and heart rate measures of attention, it was also found that long lookers spend more time in attention termination than short lookers (Colombo et al., 2001). Attention termination refers to periods when the infant is inattentive yet continues to look at the stimulus (Richards & Casey, 1991). However, conclusions regarding cortical sources of individual differences in infant visual attention remain speculative. Research with older children indicates that between 7 and 14 years of age, children's processing of global and local information becomes lateralized with global information processed primarily in the right hemisphere and local information processed primarily in the left hemisphere (Mondloch, Geldart, Maurer, & de Schonen, 2003; Moses et al., 2002). Future studies should apply cortical source localization techniques to determine if short and long lookers utilize different cortical regions when processing and encoding visual stimuli.

Conclusion

This study provides support at the neural level for differences in processing strategy based on looker type. Short lookers showed significant differences in Nc and LSW amplitude between novel-global and familiar stimuli, whereas long lookers showed significant differences in LSW amplitude between novel-local and familiar stimuli. These findings indicate the long lookers show local-precedence effects and process hierarchical patterns composed of many elements in an immature and inefficient manner, whereas short lookers show global-precedence effects similar to older children adults (Kimchi, 1990, 1992; Mondloch et al., 2003; Poirel et al., 2008).

These are the first set of findings to provide insight into processing strategies used by short- and long-looking infants at the neural level. Differential ERP responding based on stimulus type was found across looker type groups with a relatively short familiarization time of 20 s. Behavioral studies using similar stimuli have required 40–50 s of familiarization time for long lookers to demonstrate evidence of discrimination (e.g., Colombo et al., 1995). Thus, the ERP provides a valuable and efficient tool that researchers should continue to utilize to investigate individual differences and developmental processes in early cognitive development. In the current study, we examined global and local processing of hierarchical patterns composed of many elements. Global and local processing of hierarchical patterns composed of few elements remains relatively unexplored in the infant literature. Further work is needed to systematically examine various factors that may influence developmental change in global and local processing in infancy.

References

Ackles, P. K., & Cook, K. G. (2007). Attention or memory? Effects of familiarity and novelty on the Nc component in event-related brain potentials in six-month-old infants. *International Journal of Neuroscience, 117*, 837–867.

Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development from infancy. *Child Development, 57*, 251–274.

Bouvet, L., Rousset, S., Valdois, S., & Donnadieu, S. (2011). Global precedence effect in audition and vision: Evidence for similar cognitive styles across modalities. *Acta Psychologica, 138*, 329–335.

Carver, L. J., Bauer, P. J., & Nelson, C. A. (2000). Associations between infant brain activity and recall memory. *Developmental Science, 3*, 234–246.

Casey, B. J., Trainor, R., Giedd, J., Vauss, Y., Vaituzis, C. K., Hamburger, S., et al. (1997). The role of anterior cingulate in automatic and controlled processes: A developmental neuroanatomical study. *Developmental Psychobiology, 30*, 61–69.

Cohen, R. A. (1993). *The neuropsychology of attention*. New York: Plenum.

Colombo, J. (1993). *Infant cognition: Predicting later intellectual functioning*. Newbury Park, CA: Sage.

Colombo, J. (1995). On the neural mechanisms underlying developmental and individual differences in visual fixation in infancy: Two hypotheses. *Developmental Review, 15*, 97–135.

Colombo, J., Freeseman, L. J., Coldren, J. T., & Frick, J. E. (1995). Individual differences in infant fixation duration: Dominance of global versus local stimulus properties. *Cognitive Development, 10*, 271–285.

Colombo, J., Frick, J. E., Ryther, J. S., & Gifford, J. J. (1996). Four-month-olds' recognition of complementary-contour forms. *Infant Behavior and Development, 19*, 113–119.

Colombo, J., & Mitchell, D. W. (1990). Individual and developmental differences in infant visual attention. In

J. Colombo & J. W. Fagen (Eds.), *Individual differences in infancy* (pp. 193–227). Hillsdale, NJ: Erlbaum.

Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual differences in infant attention: Are short lookers faster processors or feature processors? *Child Development, 62*, 1247–1257.

Colombo, J., Mitchell, D. W., & Horowitz, F. (1988). Infant visual attention in the paired-comparison paradigm: Test-retest and attention-performance relations. *Child Development, 77*, 680–695.

Colombo, J., Mitchell, D. W., O'Brien, M., & Horowitz, F. (1987). Stability of visual habituation during the first year of life. *Child Development, 58*, 474–487.

Colombo, J., Richman, W. A., Shaddy, D. J., Greenhoot, A. F., & Maikranz, J. M. (2001). Heart rate-defined phases of attention, look duration, and infant performance in the paired-comparison paradigm. *Child Development, 72*, 1605–1616.

Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' attention to patterned stimuli: Developmental change from 3 to 12 months of age. *Child Development, 77*, 680–695.

Courchesne, E., Ganz, L., & Norcia, A. M. (1981). Event-related brain potentials to human faces in infants. *Child Development, 52*, 804–811.

De Haan, M. (2007). Visual attention and recognition memory in infancy. In M. de Haan (Ed.), *Infant EEG and event-related potentials* (pp. 101–144). New York: Psychology Press.

De Haan, M., & Nelson, C. A. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. *Child Development, 68*, 187–210.

DeBoer, T., Scott, L., & Nelson, C. A. (2007). Methods for acquiring and analyzing event-related potentials. In M. de Haan (Ed.), *Infant EEG and event-related potentials* (pp. 5–144). New York: Psychology Press.

Dukette, D., & Stiles, J. (2001). The effects of stimulus density on children's analysis of hierarchical patterns. *Developmental Science, 4*, 233–251.

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.), *Handbook of psychophysiology* (pp. 53–84). New York: Cambridge University Press.

Fagan, J. (1984). The intelligent infant: Theoretical implications. *Intelligence, 8*, 1–9.

Freeseman, L. J., Colombo, J., & Coldren, J. T. (1993). Individual differences in infant visual attention: Four-month-olds' discrimination and generalization of global and local stimulus properties. *Child Development, 64*, 1191–1203.

Frick, J. E., & Colombo, J. (1996). Individual differences in infant visual attention: Recognition of degraded visual forms by four-month-olds. *Child Development, 67*, 188-204.

Frick, J. E., Colombo, J., & Allen, J. R. (2000). Temporal sequence of global-local processing in 3-month-old infants. *Infancy, 1*, 375-386.

Frick, J. E., Colombo, J., & Saxon, T. F. (1999). Individual and developmental differences in disengagement of fixation in early infancy. *Child Development, 70*, 537–548.

Ghim, H.-R., & Eimas, P. (1988). Global and local processing by 3- and 4-month-old infants. *Perception & Psychophysics, 43*, 165–171.

Han, S., Fan, S., Chen, L., & Zhuo, Y. (1997). On the different processing of wholes and parts: A psychophysiological study. *Journal of Cognitive Neuroscience, 9*, 686–697.

Han, S., Yund, E. W., & Woods, D. L. (2003). An ERP study of the global precedence effect: The role of spatial frequency. *Clinical Neurophysiology, 114*, 1850–1865.

Hoffman, J. E. (1980). Interaction between global and local levels of a form. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 222–234.

Karrer, R., & Monti, L. A. (1995). Event-related potentials of 4-7 week-old infants in a visual recognition memory task. *Electroencephalography and Clinical Neurophysiology, 94*, 414–424.

Kimchi, R. (1983). Perceptual organization of visual patterns (Doctoral dissertation, University of California, Berkeley, 1982). *Dissertation Abstracts International, 44*, 345B.

Kimchi, R. (1988). Selective attention to global and local levels in the comparison of hierarchical patterns. *Perception and Psychophysics, 43*, 189–198.

Kimchi, R. (1990). Children's perceptual organization of hierarchical patterns. *European Journal of Cognitive Psychology, 2*, 133–149.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112*, 24–38.

Kimchi, R., Hadad, B., Behrmann, M., & Palmer, S. E. (2005). Microgenesis and ontogenesis of perceptual organization: Evidence from global and local processing of hierarchical patterns. *Psychological Science, 16*, 282–290.

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 521–535.

List, A., Justus, T., Robertson, L. C., & Bentin, S. (2007). A mismatch negativity study of local-global auditory processing. *Brain Research, 1153*, 122–133.

Macchi Cassia, V., & Simion, F. (2002). Individual differences in object-examining duration: do they reflect the use of different encoding strategies? *Cognitive Development, 17*, 119–1234.

Martin, M. (1979). Local and global processing: The role of sparsity. *Memory & Cognition, 7*, 476–484.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science, 6*, 255–260.

Miller, J. (1981). Global precedence in attention and decision. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 1161–1174.

Mondloch, C. J., Geldart, S., Maurer, D., & de Schonen, S. (2003). Developmental changes in processing of hierarchical shapes continue into adolescence. *Journal of Experimental Child Psychology*, *84*, 20–40.

Moses, P., Roe, K., Buxton, R. B., Wong, E. C., Frank, L. R., & Stiles, J. (2002). Functional MRI of global and local processing in children. *NeuroImage*, *16*, 415–424.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research*, *43*, 1–32.

Navon, D., & Norman, J. (1983). Does global precedence really depend on visual angle? *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 955–965.

Nelson, C. A., & Collins, P. F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*, 50–58.

Nelson, C. A., & Collins, P. F. (1992). Neural and behavioral correlates of visual recognition memory in 4- and 8-month-old infants. *Brain & Cognition*, *19*, 105–121.

Nelson, C. A., & Dukette, D. (1998). A cognitive neuroscience perspective on the relation between attention and memory development. In J. E. Richards (Ed.), *Cognitive neuroscience of attention: A developmental perspective* (pp. 327–362). Mahwah, NJ: Erlbaum.

Nikkel, L., & Karrer, R. (1994). Differential effects of experience on the ERP and behaviour of 6-month-old infants: Trend during repeated stimulus presentations. *Developmental Neuropsychology*, *10*, 1–11.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neuropsychology*, *72*, 184–187.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127–152.

Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest: Developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental Psychology*, *44*, 245–253.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, *112*, 516–540.

Posner, M. I., & Petersen, S. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*, 59–66.

Reynolds, G. D., Courage, M. L., & Richards, J. E. (2010). Infant attention and visual preferences: Converging evidence from behavior, event-related potentials, and cortical source localization. *Developmental Psychology*, *41*, 598–615.

Reynolds, G. D., Courage, M. L., & Richards, J. E. (in press). The development of attention. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology*. Oxford, UK: Oxford University Press.

Reynolds, G. D., & Guy, M. W. (2012). Brain-behavior relations in infancy: Intregrative approaches to examining infant looking behavior and event-related potentials. *Developmental Neuropsychology*, *37*, 210–225.

Reynolds, G. D., Guy, M. W., & Zhang, D. (2011). Neural correlates of individual differences in infant visual attention and recognition memory. *Infancy*, *16*, 368–391.

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy: An event-related potential and cortical source localization study. *Developmental Psychology*, *41*, 598–615.

Reynolds, G. D., & Richards, J. E. (2009). Cortical source localization of infant cognition. *Developmental Neuropsychology*, *34*, 312–329.

Reynolds, G. D., Zhang, D., & Guy, M. W. (in press). Infant attention to dynamic audiovisual stimuli: Look duration from 3 to 9 months of age. *Infancy*.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: An ERP study. *Developmental Science*, *6*, 312–328.

Richards, J. E., & Casey, B. J. (1991). Heart rate variability during attention phases in young infants. *Psychophysiology*, *28*, 43–53.

Rose, D., Slater, A., & Perry, H. (1986). Prediction of childhood intelligence from habituation in early infancy. *Intelligence*, *10*, 251–263.

Rugg, M. D., Fletcher, P. C., Chua, P. M.-L., & Dolan, R. J. (1999). The role of prefrontal cortex in recognition memory for source: An fMRI study. *NeuroImage*, *10*, 520–529.

Sanders, L. D., & Poeppel, D. (2007). Local and global auditory processing: Behavioral and ERP evidence. *Neuropsychologia*, *45*, 1172–1186.

Snyder, K. A. (2010). Neural correlates of encoding predict infants' memory in the paired-comparison procedure. *Infancy*, *15*, 270–299.

Snyder, K., Webb, S. J., & Nelson, C. A. (2002). Theoretical and methodological implications of variability in infant brain response during a recognition memory paradigm. *Infant Behavior and Development*, *25*, 466–494.

Srinivasan, R., Tucker, D. M., & Murias, M. (1998). Estimating the spatial Nyquist of the human EEG. *Behavioral Research Methods, Instruments, & Computers*, *30*, 8–19.

Wiebe, S. A., Cheatham, C. L., Lukowski, A. F., Haight, J. C., Muehleck, A. J., & Bauer, P. J. (2006). Infants' ERP responses to novel and familiar stimuli change over time: Implications for novelty detection and memory. *Infancy*, *9*, 21–44.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1407–1424

The Moving Window Technique: A Window Into Developmental Changes in Attention During Facial Emotion Recognition

Elina Birmingham
Simon Fraser University

Grace Iarocci
Simon Fraser University

Daniel Smilek
University of Waterloo

Tamara Meixner
University of Victoria

Christopher Kanan
University of California, San Diego

James W. Tanaka
University of Victoria

The strategies children employ to selectively attend to different parts of the face may reflect important developmental changes in facial emotion recognition. Using the Moving Window Technique (MWT), children aged 5–12 years and adults (*N* = 129) explored faces with a mouse-controlled window in an emotion recognition task. An age-related increase in attention to the left eye emerged at age 11–12 years and reached significance in adulthood. This left-eye bias is consistent with previous eye tracking research and findings of a perceptual bias for the left side of faces. These results suggest that a strategic attentional bias to the left eye begins to emerge at age 11–12 years and is likely established sometime in adolescence.

Development of Facial Emotion Recognition (FER)

Facial expressions play a crucial role in regulating everyday social interactions. Through their facial expressions, caregivers convey emotional signals to the developing child (Sorce, Emde, Campos, & Klinnert, 1985). Given their important communicative function, it is not surprising that accurate identification of emotional expressions is an integral part of social-emotional interaction and development (Izard et al., 2001). This is evidenced early in life with the formation of attachment relationships (Steele, Steele, & Croft, 2008), and in nonverbal communication (Hall & Bernieri, 2001) and later with empathizing and predicting others' behavior (Eggum et al., 2011). By adulthood, recognizing facial expressions of emotion is ubiquitous. On a daily basis, we seamlessly decode a large repertoire of subtle and dynamically changing expressions on people's faces. This skill is critical for successful social interaction, by allowing us to make inferences about other people's internal states and to change our own behavior accordingly (for reviews, see Adolphs, 2002; Adolphs & Birmingham, 2011).

Facial emotion recognition is a complex process that develops over an extended period of time in childhood and adolescence (Herba, Landau, Russell, Ecker, & Phillips, 2006; Mondloch, Geldart, Maurer, & Le Grand, 2003; Montirosso, Peverelli, Frigerio, Crespi, & Borgatti, 2008), and even into adulthood (Rump, Giovannelli, Minshew, & Strauss, 2009). Methodological differences between studies may contribute to discrepancies in the reported developmental trajectories, with some studies requiring participants to match expressions (Herba et al., 2006; Mondloch et al., 2003), others requiring labeling expressions (Montirosso et al., 2008; Rump et al., 2009), and others using forced choice designs (Thomas, De Bellis, Graham, & LaBar, 2007). Other methodological differences such as stimulus

---

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Michael Smith Foundation for Health Research (MSFHR), and the Social Sciences and Humanities Research Council of Canada (SSHRC). The authors thank Bernardus Larryant for image processing, Dr. Tom Foulsham for assistance with analysis, and all the volunteers who assisted with data collection at Face Camp at the University of Victoria. They also thank the participants and their families who made this research possible.

Correspondence concerning this article should be addressed to Elina Birmingham, Faculty of Education, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6. Electronic mail may be sent to ebirming@sfu.ca.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0021
DOI: 10.1111/cdev.12039

set (e.g., "Ekman faces"; Ekman & Friesen, 1976 vs. other stimulus sets), stimulus presentation (static vs. dynamic), and emotions tested, may all contribute to inconsistent findings.

In addition to a general increase in FER ability with age, developmental improvements are not uniform across emotion type. Expressions of happiness are correctly categorized earlier than those of fear, anger, and disgust (Boyatzis, Chazan, & Ting, 1993; Gosselin & Larocque, 2000). Vicari, Reilly, Pasqualetti, Vizzotto, and Caltagirone (2000) found no developmental improvements in school-aged children (aged 5–6, 7–8, and 9–10 years) for labeling happy expressions, for which performance was at or near ceiling. Sharper rises in performance with age are typically found for fear, disgust, and anger, but again these patterns seem to depend on the task (Montirosso et al., 2008; Vicari et al., 2000).

Attentional Strategies Involved in FER

The emergence of more skillful FER has been associated with improved cognitive functioning (Buitelaar, van der Wees, Swaab-Barneveld, & van der Gaag, 1999). Specifically, age-related improvements in the use of visual selective attention strategies may be particularly relevant to judgments of facial expression (Dadds et al., 2006). Research on the facial features attended to by adults during FER have shown promise in pinpointing mature attentional strategies for decoding emotions. Thus, the next step is to examine how attentional strategies develop to support successful FER.

A growing body of research using eye tracking has examined how adults look at faces. The findings from these studies are mixed. On the one hand, there is considerable evidence that adults show a fundamental bias to look at the eyes of faces across a variety of tasks and settings (e.g., Birmingham, Bischof, & Kingstone, 2008), including FER (e.g., Neumann, Spezio, Piven, & Adolphs, 2006; Pelphrey et al., 2002). For instance, Hall, Hutton, and Morgan (2010) found that adults (especially women) committed more fixation time to eyes than mouths during FER and that fixation time on the eyes was positively associated with accuracy of emotion judgment. Thus, there appears to be a default bias to attend to the eyes that extends to FER tasks.

On the other hand, there is some evidence to suggest that adult observers change their attentional strategy to collect optimal perceptual information for a given emotion. This is predicted by research examining the diagnostic information for decoding emotions, indicating that for certain emotions (fear and anger), the eyes are most informative, and for other emotions (happy, disgust, surprise) the mouth is more informative (Smith, Cottrell, Gosselin, & Schyns, 2005; Tanaka, Kaiser, Butler, & Le Grand, 2012). Consistent with these findings, adults are better at recognizing emotions from isolated features when the feature presented is the optimal one (Calder, Keane, Young, & Dean, 2000; Wallace, Coleman, & Bailey, 2008). Thus, it is not surprising that some eye tracking research has found that adult observers look more at the eyes when viewing facial expressions of fear than when viewing other expressions (Corden, Chilvers, & Skuse, 2008). Similarly, failing to fixate the eyes drastically reduces the ability to recognize facial expressions of fear (Adolphs et al., 2005).

However, overall the evidence from eye tracking studies is equivocal as to whether adults employ attentional strategies consistent with obtaining optimal perceptual information. For instance, Hall et al. (2010) found the expected increase in fixations to the eyes for fearful faces relative to happy and disgusted faces, but unexpectedly found that more fixations were made to the mouths of angry and fearful faces relative to disgusted faces. Other studies examining eye and mouth fixations as a function of emotion have revealed inconsistent or inconclusive results, with some data difficult to interpret because no statistical comparison was provided between the different emotions (e.g., Pelphrey et al., 2002; Sullivan, Ruffman, & Hutton, 2007).

The existing research on scan patterns in children and infants has also produced mixed results. Infants have been found to preferentially scan a face's eyes by as early as 2 months of age (Haith, Bergman, & Moore, 1977). After 4 months, there is inconsistency in the literature, with some research showing a bias to look at eyes that is similar in magnitude to adults (Hunnius, de Wit, Vrins, & von Hofsten, 2011), and other research demonstrating the emergence of a preference to look at the mouth that increases with age (Hunnius & Geuze, 2004). Other work has shown a preference for mouth over eyes in children aged 1–7 years, with a reversal of this pattern in adulthood (Nakano et al., 2010); yet other studies report an equal distribution of fixations between eyes and mouths (Amso, Fitzgerald, Davidow, Gilhooly, & Tottenham, 2010; Merin, Young, Ozonoff, & Rogers, 2007). Few studies have examined scanning patterns of infants or children while viewing emotional faces, and again these have produced inconclusive or conflicting results (Bal et al., 2010; de Wit, Falck-Ytter, & von Hofsten, 2008; Hunnius et al., 2011). Although stimulus and task factors may account for some of the differences in findings (e.g., static vs. dynamic faces; passive vs. active tasks), we now turn to an alternative possibility.

Measuring Attentional Strategies During Emotion Recognition With Eye Tracking

Findings from FER studies may be equivocal due to the method used to measure attention. Many researchers use eye tracking technology under the assumption that where an individual looks is an index of where they are allocating their attention (Henderson, 2003). Although eye tracking is a useful measure of viewing behavior, the lack of sensitivity of the studies reviewed above may be attributed to the fact that people do not necessarily commit their attention to where they are looking (e.g., Klein & Pontefract, 1994). For example, an individual may fixate the nose of a face while at the same time attending to the eyes. Indeed, previous research has demonstrated this dissociation between fixation position and information use. Caldara, Zhou, and Miellet (2010) found that under normal viewing conditions some observers predominantly fixated the nose; however, when viewing was restricted to a small *gaze-contingent* window, it became clear that these same observers relied on information from the eyes to recognize faces. Thus, while researchers studying FER have assumed that eye position and the focus of attention are one and the same, the fact that *overt orienting* (orienting accompanied by eye or head movements) and *covert orienting* (orienting without eye or head movements) can be dissociated means that eye position may not be a reliable indicator of where someone is attending.

A further issue with eye tracking is the evidence that aspects of face perception occur rapidly, within one to two fixations. Hsiao and Cottrell (2008) determined the number of fixations necessary to accurately identify a face. When the number of possible fixations was manipulated, results indicated that within the first fixation participants could recognize faces at above-chance levels. Two fixations were ideal, resulting in optimal performance; no increases in performance were observed with more than two fixations. Adult response times (RTs) for FER are also well under 1 s (De Sonneville et al., 2002). Yet, in a typical eye tracking study, the face stimulus is displayed for a fixed duration (e.g., anywhere from 2.5 s to 8 s), allowing participants to freely view the image until the end of the display period. Thus, it is possible that much of the eye movement data that are being recorded and analyzed in these studies occurs after the emotion decision has been made. This approach may severely limit what we can learn from eye tracking studies examining attention to diagnostic information during emotion recognition.

The Moving Window Technique (MWT)

In this study we use an alternate method for measuring attentional strategies across development, the MWT, capable of providing a direct measure of attention. In the MWT, participants are presented with a blurred image of a face and a small movable window that removes the blur, allowing detailed inspection within the window. The position of the window is fully controlled by the participant using a hand-held mouse. Trials are terminated by the user, which both provides a measure of RT and limits the inclusion of task-irrelevant data. Because the window provides more detailed information to the visual system, it measures directly the location of participant's attention, providing a distinct advantage over eye tracking. The MWT is an extension of the *Restricted Focus Viewer*, which was developed and validated in studies of diagram interpretation (e.g., Jansen, Blackwell, & Marriott, 2003). Variants of this technique have been used to study the processing of objects within the focus of attention (e.g., Smilek, Frischen, Reynolds, Gerritsen, & Eastwood, 2007) and in combination with eye tracking to examine attentional allocation during scene viewing (Dalrymple, Birmingham, Bischof, Barton, & Kingstone, 2011). The MWT offers similar advantages to the *gaze-contingent paradigm*, in which the position of the window is linked to the observer's eye movements, with information outside the window masked or degraded (e.g., Caldara et al., 2010; Kennedy & Adolphs, 2010; McConkie & Rayner, 1975). However, the gaze-contingent technique eliminates the normal use of peripheral vision to guide eye movements, creating an unnatural viewing situation and abnormally shortened saccades (Foulsham, Teszka, & Kingstone, 2011; Loschky & McConkie, 2002). In contrast, in the MWT the window is not linked to fixation position, but is movable by the mouse, allowing for a more natural viewing experience relative to gaze-contingent methods (Dalrymple et al., 2011).

Here we present a novel adaptation of the mouse-contingent restricted viewing approach to study how children explore faces. The use of a mouse-controlled window is simple enough for young children to complete (Donker & Reitsma,

2007) and may confer the advantages of being cheaper and easier to administer than some eye trackers, especially as no calibration is required for the MWT, one of the most common sources of error in eye tracking studies (Jansen et al., 2003). In this study, we used the MWT to investigate the development of attentional strategies involved in FER. Four basic expressions (happy, disgusted, angry, fearful) were included because children as young as 3½ years old are capable of recognizing these expressions at above-chance accuracy levels (Boyatzis et al., 1993). Also, including two of each of the so-called "top" emotions (diagnostic information in the eyes: *angry* and *fearful*) and "bottom" emotions (diagnostic information in the mouth: *happy* and *disgusted*), allowed us to investigate how attentional strategies change as a function of both age and emotion type.

Method

Participants

Five age groups were tested in this experiment ($N = 129$). Twenty-two 5- to 6-year-olds (11 male participants, $M_{\text{age}} = 5.5$ years), twenty-eight 7- to 8-year-olds (16 male participants, $M_{\text{age}} = 7.4$ years), thirty 9- to 10-year-olds (13 male participants, $M_{\text{age}} = 9.7$ years), twenty-four 11- to 12-year-olds (15 male participants, $M_{\text{age}} = 11.3$ years), and 25 undergraduate students (4 male participants, $M_{\text{age}} = 21.5$ years) participated in this study. Participants aged 5–12 years were tested during a free 1-day camp held at the University of Victoria; adult participants were students from the University of Victoria who completed the experiment for course credit. All participants had normal or corrected-to-normal vision. The study protocol was conducted in accordance with the ethical guidelines of the University of Victoria.

Stimuli

Stimuli were taken from the NimStim set of facial expressions (Tottenham et al., 2009; see Figure 1a for examples). Faces were modified in Adobe Photoshop CS4 (San Jose, CA) by converting each image to grayscale and blurring using a Gaussian filter with a standard deviation equal to 40 pixels, corresponding to approximately 12% of the face width. Blurring to this extent ensured that the facial expression could not be identified without exploring the face with the window (see Figure 1b). All faces were displayed on a white canvas measuring 1280 × 800 pixels in size, subtending 27 × 17 degrees of visual angle (°) at a viewing distance of approximately 60 cm. After consulting the available validation data for the NimStim faces (Tottenham et al., 2009, Supplementary Data), 10 exemplars (5 male exemplars, 5 female exemplars) were chosen for each of the four expression types (40 exemplars total) on the basis that they were unambiguous and yielded high identification accuracy from an independent sample of judges, *mean accuracy percentages* (*SD*): happy = 0.98 (0.02), angry = 0.91 (0.12), disgusted = 0.94 (0.04), fearful = 0.79 (0.07). Face stimuli included European American and Latino American ethnicities, and were posed with an open mouth to ensure that the task would not be too difficult for the younger children.

Procedure

Moving Window Condition

Participants were tested on a 13-in. MacBook Pro running Snow Leopard (10.6.3). Participants were tested individually with the aid of a research assistant (RA) who was blind to the hypotheses of the study. Participants sat comfortably in front of the computer with an unobstructed view of the screen and access to an external hand-held mouse. The RA sat next to the participant throughout the experiment.

Before beginning the experiment, the computer presented four faces labeled as "happy," "angry," "grossed out" (a more familiar term for young children than "disgusted"), and "scared." These prototype faces (not used in the experiment) were reviewed by the RA and the participant to familiarize him or her with the four emotions. Following this, a screen appeared with the instructions: "Can you tell me what the face is feeling? A blurry face is going to appear on the screen, and you will be able to explore the face through a window on the screen. You can reveal different parts of the face by moving the window around the screen with the mouse. You'll have 15 s to explore. When you know the feeling, say it out loud! Can you tell me the feeling before the face disappears?" Then, four practice trials were presented to ensure that the participant understood the task and practiced using the mouse to explore the faces.

There were a total of 40 experimental trials. Presentation order was randomized across participants. Each trial began with a prompt ("Ready?") appearing in the center of the screen to focus the participant's attention; the RA then pressed the space bar

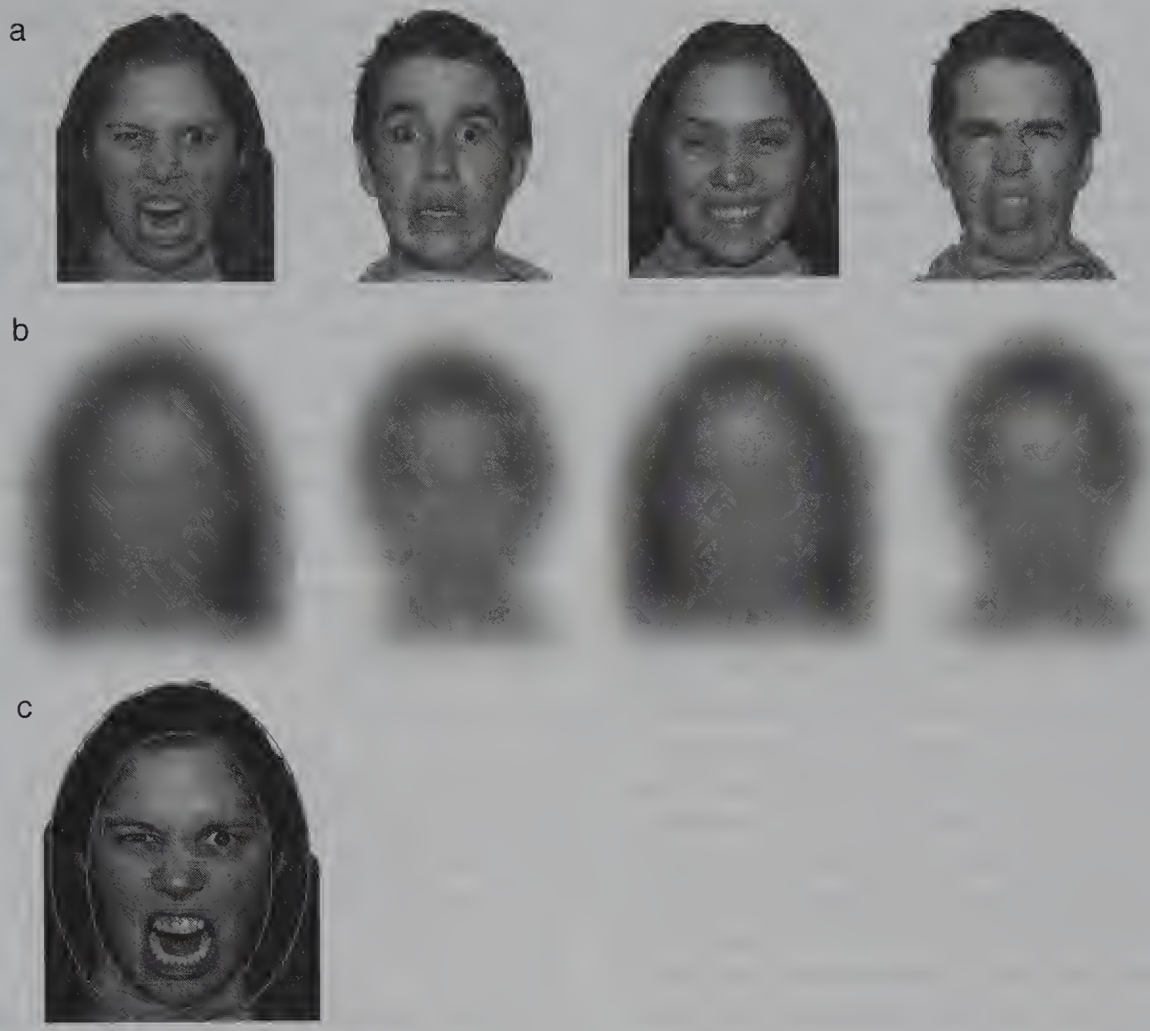


Figure 1. (a) Sample faces expressing anger, fear, happiness, and disgust. (b) Sample faces after being blurred using a Gaussian filter with a radius of 40 pixels. (c) The following regions of interest were created for each individual stimulus: left eye, right eye, nose, mouth, face remainder, and head (hairline/ears).

revealing a blurred face stimulus in the middle of the screen. The mouse was linked to the center of a 100 × 100 pixel (2 × 2°) window, an area spanning approximately 4% of the face. The window removed the Gaussian blur, revealing detailed visual information within the window. The start position of the window was always the center of the screen. Participants then explored the image using the mouse to determine the emotion. To reduce memory load, the emotion labels were presented on the screen. Although no feedback was given to participants on their responses, to motivate and maintain interest we pseudo-randomly presented images of popular cartoon characters delivering positive messages (e.g., an image of Shrek saying, "Fantastic work!").

The participant was told to verbally respond, at which time the RA immediately pressed the space bar to terminate the trial and entered the response using one of four appropriately labeled keys on the keyboard. If the participant did not respond within 15 s, the trial timed out and a response screen appeared, prompting the participant to make his or her best guess. After the RA entered the participant's response, the "Ready?" prompt for the next trial appeared. Undergraduate participants followed an identical procedure. Mouse coordinate data were sampled at a rate of 60 Hz and written to a text file.

Baseline Condition

To determine baseline accuracy of emotion identification without the requirement to explore the faces, immediately following completion of the moving window condition, participants were introduced to the baseline condition. The procedure was identical to that of the moving window task with the exception that stimuli were presented in full resolution. As our primary goal was to determine how children visually explore a novel set of expressive faces, the baseline condition was necessarily always presented after the moving window condition to prevent carryover effects.

Analysis

Moving Window Condition

Accuracy and RT. We examined accuracy and RTs as a function of both age and emotion type. These data were submitted to mixed 5 × 4 repeated measures analyses of variance (ANOVAs) in which age group (5–6, 7–8, 9–10, and 11–12 years, adult) was the between-subjects factor and emotion type (happy, disgusted, angry, fearful) was the within-subjects factor. Subsequent post hoc comparisons used Fisher's least significant difference test (LSD) method ($\alpha = .05$) to further explore all significant effects. An exploratory analysis of error patterns was also conducted.

Exploration of regions of interest (ROIs). The following mutually exclusive ROIs were defined: left eye, right eye, mouth, nose, face remainder (face minus eyes, nose, and mouth), and head (hairline and ears; see Figure 1c). To control for differences in overall exploration time across the groups and emotion types, we computed a *proportion of time* spent in each ROI, a value that reflects both the time and the amount of ROI information (pixels) revealed by the window during exploration. The following steps were involved in computing these proportion values: (a) we measured the pixel overlap between the window and each ROI at each time sample and computed a Cumulative Area × Time value for each ROI on each trial; (b) we normalized this value by the area of the ROI to correct for the fact that bigger ROIs would, by definition, have greater overlap with the window than smaller ROIs; and (c) we computed a proportion based on these normalized values [e.g., time proportion for left eye = eyeL/(eyeL + eyeR + nose + mouth + face + head)]. To analyze the time proportion data we conducted the repeated measures ANOVA with age group as the between-subjects factor, and emotion type and ROI as the within-subjects factors. Subsequent post hoc tests used the Fisher's LSD method ($\alpha = .05$) to further explore all significant effects.

First ROIs explored. We examined the first three ROIs explored as a function of age group, collapsing over emotion type. A frequency value was computed for each ROI separately for each subject, based on whether the center of the exploration window entered an ROI. For example, if the center of the window entered the left eye first, the nose second, and the mouth third, the frequencies would be as follows, *first ROI*: eyeL = 1, eyeR = 0, nose = 0, mouth = 0, face = 0, head = 0; *second ROI*: eyeL = 0, eyeR = 0, nose = 1, mouth = 0, face = 0, head = 0; *third ROI*: eyeL = 0, eyeR = 0, nose = 0, mouth = 1, face = 0, head = 0. These values were averaged over trials and submitted to repeated measures ANOVA with age group as the between-subjects factor and ROI as within-subjects factors, followed up with post hoc pairwise comparisons.

Last ROI explored prior to response. To investigate which region of the face was attended to at the time of response, we examined the last ROI that was explored on the face before the spacebar was pressed to end the trial. Frequencies (as in the first ROIs analysis) for each ROI were computed separately for each subject and averaged across trials. The repeated measures ANOVA was conducted on this data, with age group as the between-subjects factor, and emotion type and ROI as the within-subjects factors, followed up with post hoc pairwise comparisons.

Baseline Condition

Accuracy and RT. Due to fatigue or computer error, two participants in the 5- to 6-year-old group did not complete the baseline condition. As with the moving window condition, we examined accuracy and RTs as a function of age and emotion type. These data were submitted to mixed 5 × 4 repeated measures ANOVAs in which age group was the between-subjects factor and emotion was the within-subjects factor.

Results

Moving Window Condition

Accuracy (Proportion Correct)

Figure 2a illustrates the accuracy data as a function of age group and emotion type. There was a significant main effect of age group, $F(4, 124) = 11.39$, $p < .001$. Post hoc comparisons revealed that 5- to 6-year-olds ($M = 0.84$) were the least accurate and significantly less accurate than all other age groups (7–8: $M = 0.90$; 9–10: $M = 0.92$; 11–12: $M = 0.96$; adult: $M = 0.94$; Fisher's LSD $p < .05$) and that 11- to 12-year-olds were more accurate than 7- to 8-year-olds and 9- to 10-year-olds, $p < .05$, but no different from adults, $p > .05$.

The main effect of emotion type, $F(3, 372) = 32.03$, $p < .001$, was also significant, with accuracy highest for happy faces ($M = 0.99$) followed by disgusted faces ($M = 0.93$), which were more accurately identified than both fearful ($M = 0.87$) and angry ($M = 0.86$) faces ($p < .05$). Accuracy for fearful and angry faces did not differ, $p > .05$. The interaction

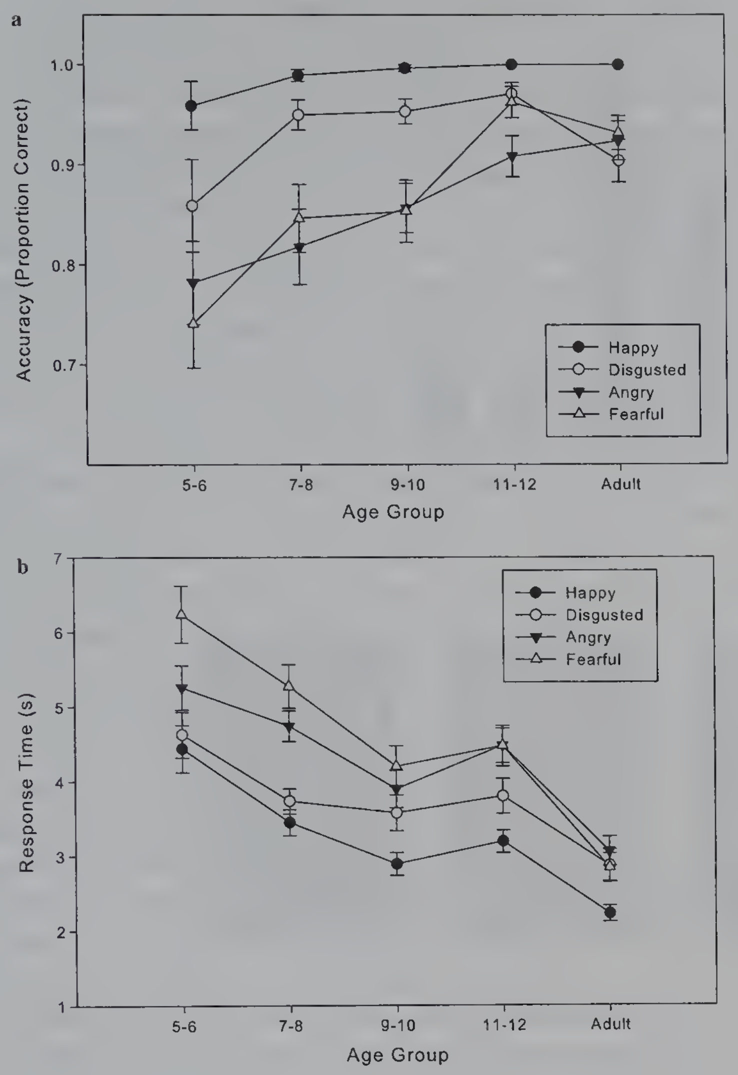


Figure 2. (a) Accuracy as a function of age group and emotion. *Age group*: Five- to 6-year-olds were the least accurate and significantly less accurate than all other age groups, $p < .05$. *Emotion type*: Accuracy was highest for happy > disgust > fear = angry, $p < .05$. *Age Group × Emotion Type*: Five- to 6-year-olds were particularly poor for fearful and angry faces. (b) Response time (RT) as a function of age group and emotion. *Age group*: Adults were faster to respond than all other age groups, $p < .05$. *Emotion type*: Responses were fastest for happy and slowest for fearful, followed by angry and disgusted faces, $p < .05$. *Age Group × Emotion Type*: Five- to 6-year-olds were especially slow to respond to fearful faces. RT, for fear in particular, tended to reduce with age. Symbols are means, and error bars are standard errors.

between age group and emotion type was also significant, $F(12, 372) = 2.37$, $p < .01$, reflecting different developmental trajectories for the emotions. Separate ANOVAs on each emotion type revealed that performance was near ceiling for happy faces, with little age-related improvement, as can be seen in Figure 2a, $F(4, 124) = 2.73$, $p < .05$; all age groups were slightly but significantly more accurate than 5- to 6-year-olds, $p < .05$. In contrast, there was a stepwise improvement in identifying disgusted faces from ages 5–6 years ($M = 0.86$) to 7–8 years ($M = 0.95$), $p < .05$, $F(4, 124) = 3.69$, $p < .01$, but no improvements thereafter. Finally, angry and fearful faces produced more systematic improvements in accuracy with age: angry, $F(4, 124) = 3.70$, $p < .01$; fearful, $F(4, 124) = 7.27$, $p < .001$, with accuracy particularly low at age 5–6 years, but improving slightly at each age. For angry faces, there was a significant improvement in accuracy by age 11–12 years ($M = 0.91$) relative to 5–6 years ($M = 0.78$) and 7–8 years ($M = 0.82$); for fearful faces, there was a significant improvement from age 5–6 ($M = 0.73$) to 7–8 ($M = 0.85$) years, and again at age 11–12 years ($M = 0.96$), relative to both 7–8 and 9–10 years, $p < .05$.

Error Patterns

An examination of error type as a function of age group and emotion revealed that 5- to 6-year-ods were more likely than all other age groups to label fearful faces as disgusted (see Table 1). This was confirmed by submitting the frequency of "disgusted" responses to an Age Group × Emotion Type ANOVA, revealing an interaction, $F(8, 248) = 2.25$, $p < .05$, Fisher's post hoc tests, $p < .05$. In addition, by submitting the frequency of "angry" responses to an Age Group × Emotion Type ANOVA, we found that 5- to 6-year-olds and adults were more likely than all other age groups to label disgusted faces as angry, $F(8, 248) = 2.48$, $p < .05$, Fisher's post hoc tests, $p < 0.05$. There were no age-related differences in the other error types, $p > .05$.

Response Time

The RT data for one 5- to 6-year-old participant was abnormally slow and therefore excluded from analysis. The main effect of age group, $F(4, 123) = 17.33$, $p < .001$ was reliable. Post hoc tests revealed that 5- to 6-year-olds ($M = 5.1$ s) responded more slowly than 7- to 8-year-olds ($M = 4.3$ s), 9- to 10-year-olds ($M = 3.6$ s), 11- to 12-year-olds ($M = 4.0$ s), and adults ($M = 2.7$ s), $p < .05$. Adults were faster than all children, $p < .05$. The main effect of emotion type, $F(3, 369) = 89.81$, $p < .001$, was also significant, with post hoc tests revealing that, consistent with the accuracy data, participants were fastest for happy faces ($M = 3.2$ s), followed by disgusted faces ($M = 3.7$ s), angry faces ($M = 4.3$ s), and slowest for fearful faces ($M = 4.6$ s; happy < disgust < angry < fearful, $p < .05$). Importantly, these results argue against a speed–accuracy trade-off. Finally, as with the accuracy data, the interaction between age group and emotion type was reliable, $F(12, 369) = 4.53$, $p < .001$, reflecting that 5- to 6-year-olds were particularly slow to recognize fearful faces

Table 1
Means of Error Types as a Function of Age Group and Emotion

Response (error type)	Age group				
	5–6	7–8	9–10	11–12	Adult
	Emotion of face				
Happy	H: —	H: —	H: —	H: —	H: —
	D: 0.01	D: 0.00	D: 0.01	D: 0.01	D: 0.00
	A: 0.02	A: 0.01	A: 0.00	A: 0.01	A: 0.01
	F: 0.02	F: 0.01	F: 0.01	F: 0.01	F: 0.00
Disgusted	H: 0.01	H: 0.00	H: 0.00	H: 0.00	H: 0.00
	D: —	D: —	D: —	D: —	D: —
	A: 0.13	A: 0.11	A: 0.08	A: 0.03	A: 0.05
	F: 0.20*	F: 0.13	F: 0.13	F: 0.03	F: 0.06
Angry	H: 0.01	H: 0.01	H: 0.01	H: 0.00	H: 0.00
	D: 0.08*	D: 0.04	D: 0.03	D: 0.02	D: 0.08*
	A: —	A: —	A: —	A: —	A: —
	F: 0.04	F: 0.01	F: 0.02	F: 0.01	F: 0.01
Fearful	H: 0.01	H: 0.01	H: 0.00	H: 0.00	H: 0.00
	D: 0.01	D: 0.01	D: 0.02	D: 0.01	D: 0.02
	A: 0.08	A: 0.06	A: 0.06	A: 0.05	A: 0.02
	F: —	F: —	F: —	F: —	F: —

Note. Five- to 6-year-olds were more likely than all other age groups to label fearful faces as disgusted. Five- to 6-year-olds and adults were more likely than all other age groups to label disgusted faces as angry. H = happy; D = disgusted; A = angry; F = fearful.
*Significantly different from the other groups at $p < .05$.

($M = 6.2$ s) and that RT for fear, in particular, tended to decrease with age (see Figure 2b).

Time Spent Exploring ROIs

We analyzed the *proportion of time* in each ROI (see Figure 3a), revealing how exploration biases change with age, taking into account developmental changes in overall exploration time. One 5- to 6-year-old participant perseverated on the right eye to the exclusion of all other regions and was thus removed from analysis. An ANOVA with age group, emotion type, and ROI as factors revealed a main effect of ROI, $F(5, 615) = 479.40$, $p < .001$, suggesting that participants spent the greatest proportion of time on the mouth ($M = 0.39$) followed by the nose ($M = 0.29$), the left eye ($M = 0.15$), the right eye ($M = 0.11$), the face ($M = 0.06$), and finally, the head ($M = 0.01$), $p < .05$. There was also a significant Age Group × ROI interaction, $F(20, 615) = 2.70$, $p < .001$, reflecting that attention to the ROIs differed as a function of age. Separate ANOVAs conducted on each ROI revealed a significant effect of age group on the proportion of time spent exploring the left eye, $F(4, 123) = 3.27$, $p < .05$, but not the right eye, $F(4, 123) = 1.26$, $p > .20$. Post hoc tests confirmed an age-related increase in exploration time on the left eye; that is, adults spent more time exploring the left eye ($M = 0.19$) than 5- to 6-year-olds ($M = 0.12$), 7- to 8-year-olds ($M = 0.12$), and 9- to 10-year-olds ($M = 0.13$), $p < .05$, but not 11- to 12-year-olds ($M = 0.16$), who did not differ significantly from younger children, $p > .05$. Importantly, we observed a bias to explore the left eye over the right eye that was significant in adulthood, Fisher's $p < .05$, but not in children, $p > .05$.

Conversely, adults spent less time on the mouth ($M = 0.34$) than younger children, $F(4, 123) = 3.33$, $p < .05$ (adults vs. 5- to 6-year-olds: $M = 0.43$, 7- to 8-year-olds: $M = 0.43$, and 9- to 10-year-olds, $M = 0.41$, $p < 0.05$), but not 11- to 12-year-olds ($M = 0.36$). Similarly, 11- to 12-year-olds explored the mouth less than 7- to 8-year-olds, $p < .05$. Finally, there was an age-related reduction in attention to the head region, $F(4, 123) = 9.52$, $p < .001$; while little exploration time was devoted to the head, younger children spent more exploration time on this region than older children, Fisher's $p < .05$. See Figure 3b for heat maps illustrating exploration biases for each age group.

The interaction between emotion type and ROI, $F(15, 1845) = 48.87$ $p < .001$ was also significant (see Figure 4). Consistent with what we expected based on previous research (Smith et al., 2005), separate ANOVAs on each ROI revealed that participants explored the left eye more on angry ($M = 0.16$) and fearful ($M = 0.17$) faces than on disgusted ($M = 0.13$) or happy ($M = 0.14$) faces, $F(3, 369) = 24.05$, $p < .001$ (Fisher's LSD $p < .05$). The same pattern was found for the right eye, $F(3, 369) = 18.83$, $p < .001$. The mouth was explored for longer on disgusted faces ($M = 0.46$) than on happy ($M = 0.39$), fearful ($M = 0.38$) and angry ($M = 0.35$) faces, $F(3, 369) = 76.60$, $p < .001$ (Fisher's LSD, $p < .05$). The nose was explored for longer on angry ($M = 0.31$) and happy ($M = 0.31$) faces than on disgusted ($M = 0.26$) and fearful ($M = 0.27$) faces, $F(3, 369) = 40.39$, $p < .001$. Finally, the face remainder was explored slightly more on angry ($M = 0.06$) faces than on disgusted faces ($M = 0.05$) and fearful faces ($M = 0.05$), $F(3, 369) = 4.02$, $p < .01$. There was no effect of emotion type for the head region.

First ROIs Explored

Separate Age Group × ROI ANOVAs on each of the first three ROIs visited revealed main effects of

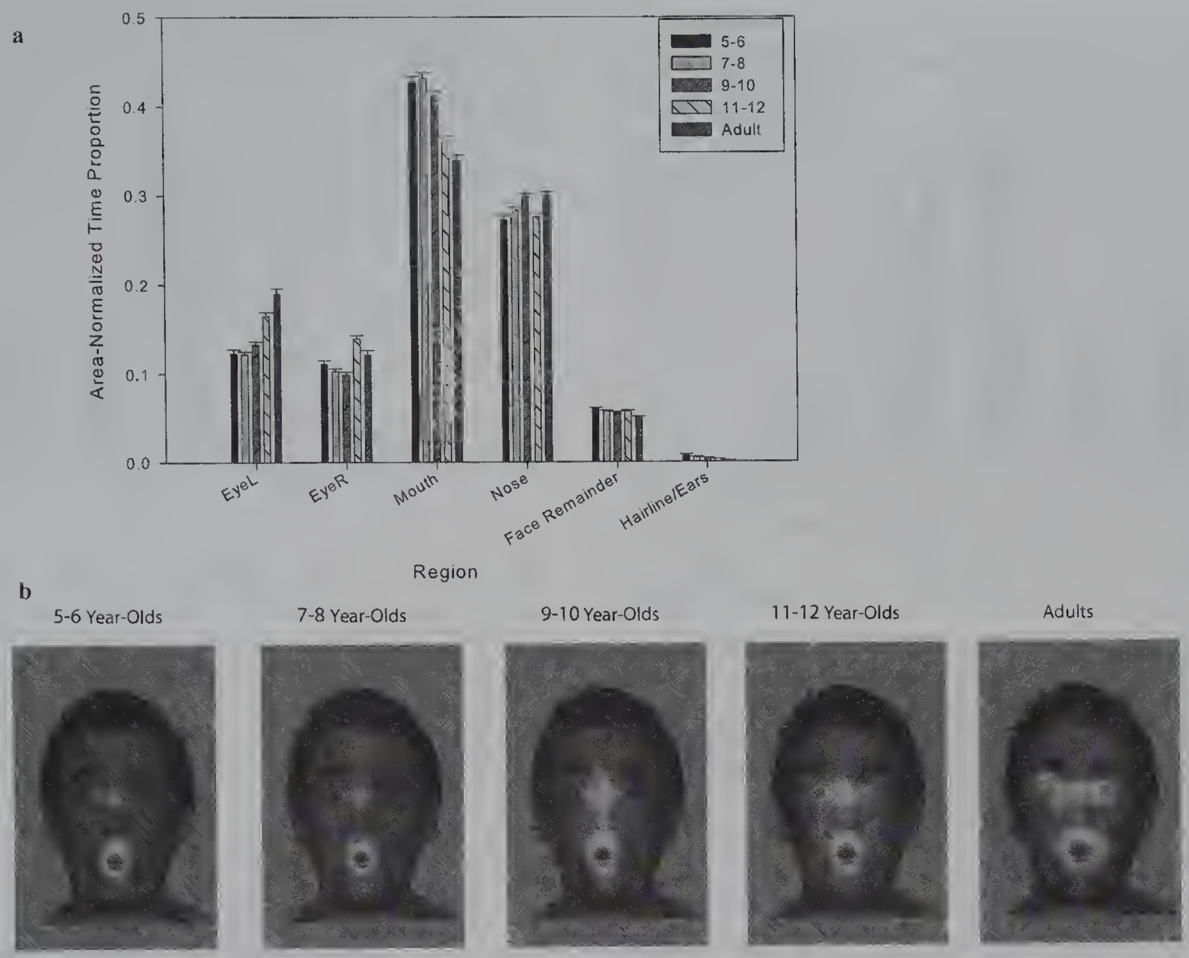


Figure 3. (a) Proportion of exploration time in each region of interest as a function of age group. There was an age-related increase in exploration time on the left eye, reaching significance in adulthood, $p < .05$. Adults spent less time on the mouth than younger children, $p < .05$. Bars are means, and error bars are standard errors. (b) Exploration heat maps showing location of exploration time as a function of age. The data for one of the fearful faces are presented. Samples were smoothed with a 48-pixel Gaussian filter. Each individual heat map is scaled from zero time (blue) to the point of maximal exploration time (red; see color version of figure online).

ROI: first ROI, $F(5,\ 615) = 190.16$, $p < .0001$; second ROI, $F(5,\ 615) = 101.35$, $p < .0001$; and third ROI, $F(5,\ 615) = 266.38$, $p < .0001$). *First ROI:* The nose was the most likely to be the very first ROI explored. This initial exploration of the nose was not unexpected because the window position at the start of each trial was at the center of the screen. After the nose, the left and right eyes were highly likely to be the very first ROI explored by participants, significantly more so than the mouth, face, and head. *Second ROI*: A different pattern emerged for the second ROI to be visited, which was again most often the nose, but this time the mouth was also highly likely to be visited, and more so than the eyes, $p < .05$. *Third ROI*: The third ROI to be visited was highly (and equally) likely to be either the mouth or nose, whereas the eyes and head were visited much less often. Importantly, there was an Age Group × ROI interaction for the first ROI, $F(20,\ 615) = 3.87$, $p < .001$, with post hoc tests confirming that adults were more likely than children to begin their exploration on the left eye, $p < .05$ (see Figure 5b). This age difference was not present for the right eye.

Last ROI Explored Prior to Response

Participants terminated their exploration most often on the mouth ($M = 0.38$) followed by the nose ($M = 0.23$), then the eyes (left: $M = 0.14$, right: $M = 0.15$), and much less often on the face ($M = 0.06$) and finally the head ($M = 0.01$), reflected by a main effect of ROI, $F(5,\ 615) = 101.90$, $p < .001$. However, an Age Group × ROI interaction, $F(20,\ 615) = 3.02$, $p < .001$, revealed that the age groups differed on where they ended their exploration (see Figure 6a). Separate ANOVAs on each ROI followed by post hoc tests confirmed that 11- to 12-year-olds were more likely to end their exploration on the left eye ($M = 0.22$) than younger age groups (5–6: $M = 0.09$, 7–8: $M = 0.10$, 9–10: $M = 0.13$), and adults were more likely to end exploration on the left eye ($M = 0.18$) than 5- to 6-year-olds and 7- to 8-year-olds, $F(4,\ 123) = 4.52$ $p < .01$. There was no

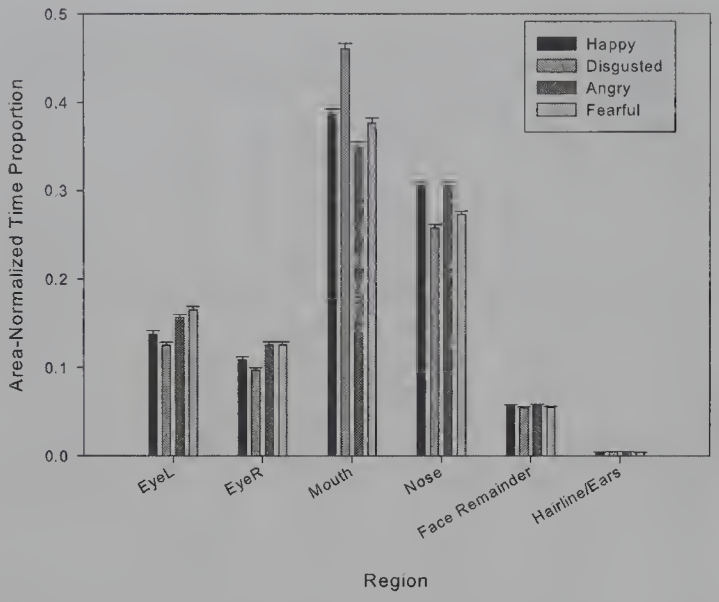


Figure 4. Proportion of exploration time in each region of interest plotted as a function of emotion type. Eyes were explored more on angry and fearful faces than on disgusted and happy faces, $p < .05$. Mouths were explored more on disgusted faces than on faces depicting other emotions, $p < .05$. Bars are means, and error bars are standard errors.

reliable effect of age group for the right eye, *F*(4, 123) = 2.13, $p = .08$. Interestingly, adults were more likely than children (except 9- to 10-year-olds) to end exploration on the nose, *F*(4, 123) = 4.71 $p < .01$. Conversely, there was a trend toward an age-related reduction in ending exploration on the mouth, *F*(4, 123) = 2.28 $p = .06$, and a highly significant reduction in ending exploration on the face remainder, *F*(4, 123) = 5.31 $p < .001$, and head, *F*(4, 123) = 5.65 $p < .001$.

The interaction between emotion type and ROI was significant, *F*(15, 1845) = 10.84, $p < .001$ (see Figure 6b). Separate ANOVAs conducted on each ROI revealed a significant effect of emotion type for the left eye, *F*(3, 369) = 4.32 $p < .01$, and right eye, *F*(3, 369) = 5.10 $p < .01$, in which participants were more likely to end on the eyes on fearful (left: $M = 0.16$, right: $M = 0.17$) and angry (left: $M = 0.16$, right: $M = 0.17$) faces than on happy (left: $M = 0.13$, right: $M = 0.13$) and disgusted (left: $M = 0.13$, right: $M = 0.12$, Fisher's LSD, $p < .05$) faces. In contrast, participants were more likely to end on the mouth on disgusted ($M = 0.46$) and happy ($M = 0.39$) faces than on angry ($M = 0.30$) and fearful ($M = 0.35$) faces, *F*(3, 369) = 24.92, $p < .001$ (Fisher's $p < .05$). Participants were more likely to end their exploration on the nose of happy faces ($M = 0.26$) than on disgusted ($M = 0.19$) and fearful ($M = 0.22$) faces, and more likely to end on the nose of angry faces ($M = 0.25$) than on disgusted faces, *F*(3, 369) = 6.76, $p < .001$. The three-way Age Group × Emotion Type × ROI interaction did not reach significance, *F*(60, 1845) = 1.14, $p > .20$.

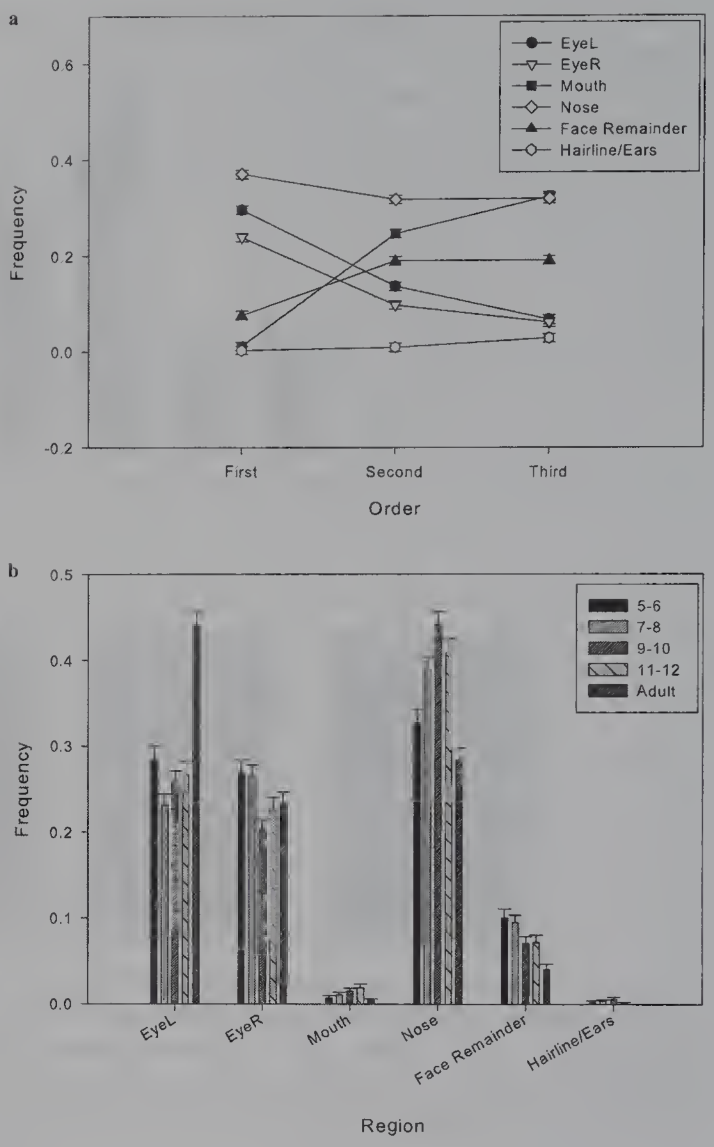


Figure 5. (a) The first three regions of interest (ROIs) to be visited. Overall, the nose was most likely to be explored within the first three visits to an ROI. However, the very first exploration was more likely to be to one of the eyes than to the mouth, $p < .05$. In contrast, the second and third visits to an ROI were more likely to be to the mouth than to either of the eyes, $p < .05$. (b) Adults were more likely than children to immediately move their mouse to the left eye (graph shows first ROI to be visited). Bars are means, and error bars are standard errors.

Correlations Between Eye Versus Mouth Exploration and Accuracy

To what extent does the strategic distribution of attention between eye and mouth regions contribute to accurate identification of emotions? Because both eye and mouth time differed as a function of age, we computed a normalized difference score of (*eyetime* − *mouthtime*)/(*eyetime* + *mouthtime*), calculated as the Cumulative Time × Area Value (for both

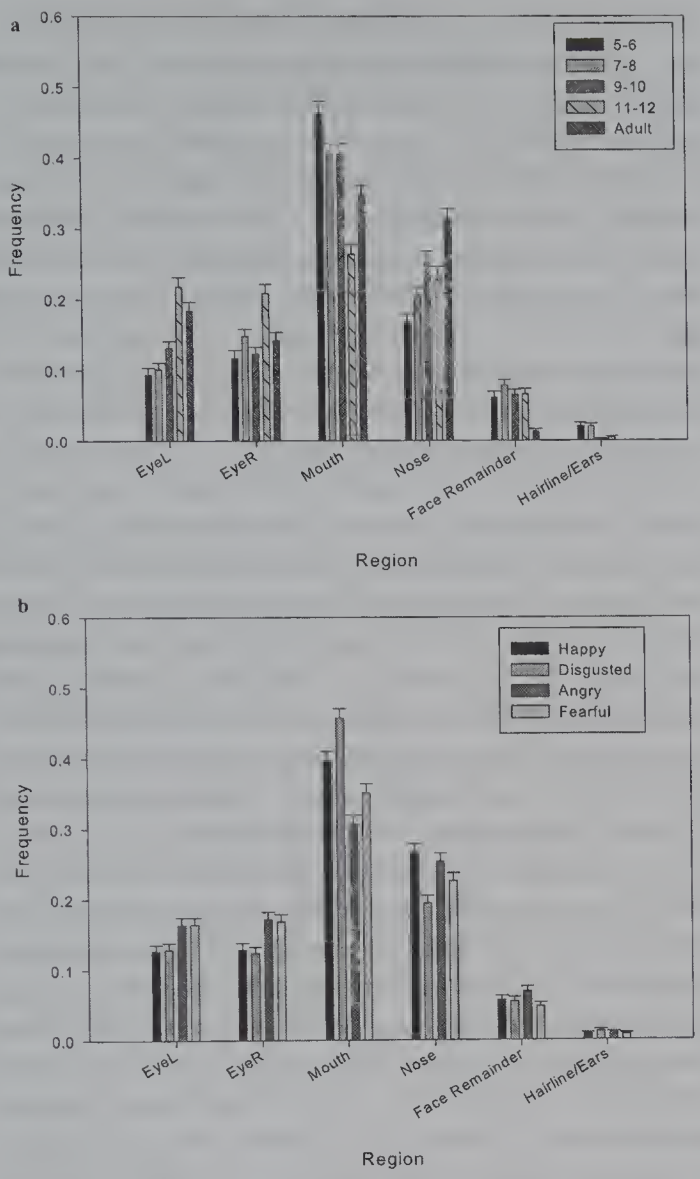


Figure 6. (a) The last region of interest to be visited before response, as a function of age. Eleven- to 12-year-olds and adults were more likely than younger children to end their exploration on the left eye than younger children, $p < .05$. Bars are means, errors bars are *SE*. (b) Participants were more likely to end on the eyes of angry and fearful faces than on happy and disgusted faces, and more likely to end on the mouths of happy and disgusted faces than on fearful and angry faces. $p < .05$ Bars are means, error bars are *SE*.

eyes) minus the Cumulative Time × Area Value for the mouth, divided by the total time spent on eyes and mouths. This was computed separately for each subject and emotion type. Thus, the more positive the normalized difference score, the more time was allocated to the eyes versus the mouth; in contrast, the more negative the difference score, the more time was allocated to the mouth versus the eyes. To determine the relation between eye exploration and accuracy, we examined partial correlations between accuracy and *eyetime* – *mouthtime*, controlling for the effect of age, separately for each emotion type.

For angry faces, we found a modest but significant positive association between *eyetime* – *mouthtime* and accuracy ($r = .26$, $p < .01$). This suggests that participants who spent more time exploring the eyes relative to the mouth were more accurate at identifying angry faces. Similarly, for fearful faces, *eyetime* – *mouthtime* and accuracy were positively correlated ($r = .26$, $p < .01$), again suggesting that exploring the eyes increased the likelihood of recognizing fearful expressions. In contrast, *eyetime* – *mouthtime* was not significantly correlated with accuracy for happy faces ($r = -.02$) but was negatively correlated with accuracy for disgusted faces ($r = -.24$, $p < .01$), suggesting that increasing exploration of the *mouth* was associated with a higher likelihood of recognizing disgusted faces.

Baseline Condition: Accuracy and RT

Accuracy (Proportion Correct)

As expected, there was a main effect of age group, $F(4, 122) = 3.80$, $p < .01$, indicating a developmental increase in accuracy. Fisher's post hoc tests revealed that 5- to 6-year-olds ($M = 0.89$) were less accurate than 9- to 10-year-olds ($M = 0.93$), 11- to 12-year-olds ($M = 0.95$), and adults ($M = 0.95$). In addition, 11- to 12-year-olds and adults were significantly more accurate than 7- to 8-year-olds ($M = 0.90$), $p < .05$. A significant effect of emotion type, $F(3, 366) = 58.79$, $p < .001$, indicated that participants were less accurate at identifying angry faces ($M = 0.81$) than all other emotions (Fisher's $p < .05$). Participants were most accurate at identifying happy faces ($M = 0.99$), followed by disgusted faces ($M = 0.97$), which were both higher than fearful faces ($M = 0.92$, Fisher's $p < .05$). Unlike in the moving window condition, the interaction between age group and emotion type did not reach significance, $p > .10$ (see Table 2).

Response Time

The ANOVA demonstrated a main effect of age group, $F(4, 122) = 23.47$, $p < .001$. Post hoc tests indicated that adults ($M = 1.8$ s) were faster than all other groups, ($p < .05$) and that 5- to 6-year-olds ($M = 3.6$ s) were slower than 9- to 10-year-olds ($M = 2.5$ s) and 11- to 12-year-olds ($M = 2.6$ s), $p < .05$, but not 7- to 8-year-olds ($M = 3.0$ s), $p > .05$. A significant effect of emotion type, $F(3, 366) = 85.44$, $p < .001$, revealed that participants were quickest to respond to happy faces ($M = 2.2$ s) followed by disgusted faces ($M = 2.6$ s). Responses to angry

Table 2
Baseline Condition: Accuracy and Response Time Data as a Function of Age Group and Emotion Type

	Emotion type			
Age group	Happy	Disgusted	Angry	Fearful
Accuracy				
5–6	0.98 (0.01)	0.95 (0.04)	0.78 (0.03)	0.86 (0.04)
7–8	1.00 (—)	0.96 (0.02)	0.75 (0.04)	0.89 (0.04)
9–10	0.99 (0.01)	0.97 (0.01)	0.85 (0.03)	0.92 (0.02)
11–12	1.00 (0.01)	0.99 (0.01)	0.83 (0.04)	0.98 (0.01)
adults	1.00 (0.01)	0.98 (0.01)	0.85 (0.03)	0.97 (0.01)
Response time (s)				
5–6	3.14 (0.22)	3.52 (0.28)	3.81 (0.30)	3.75 (0.20)
7–8	2.39 (0.17)	2.86 (0.14)	3.53 (0.20)	3.40 (0.17)
9–10	2.01 (0.10)	2.36 (0.10)	2.83 (0.14)	2.95 (0.17)
11–12	2.12 (0.11)	2.42 (0.09)	3.07 (0.10)	2.96 (0.14)
Adults	1.32 (0.04)	1.74 (0.11)	2.09 (0.11)	1.89 (0.10)

Note. Data are presented as means (standard errors).

($M = 3.1$ s) and fearful ($M = 3.0$ s) faces were significantly slower than to both happy and disgusted faces, $p < .05$. Angry and fearful faces did not differ, $p > .05$. There was no interaction between age group and emotion type, $p > .10$ (see Table 2).

Discussion

Using a novel paradigm (MWT) to measure the allocation of attention during FER, we found that with age there was a specific increase in attention to the left eye that reached significance in adulthood. Conversely, there was an age-related reduction in time spent exploring the mouth, emerging at age 11–12 years and reducing further in adulthood. Thus, age 11–12 years may mark the beginning of a developmental shift in FER toward relying more on the eyes, and less on the mouth, relative to younger children. This suggestion is bolstered by the last ROI data, in which 11- to 12-year-olds were more likely to end their exploration on the left eye (i.e., at the time of response) compared to younger children. Thus, it is perhaps not surprising that age-related improvements in accuracy and RT were most evident for the expression of fear, which previous research has found to rely on the eye region for accurate recognition (Adolphs et al., 2005; Smith et al., 2005). Thus, we suggest that at age 11–12 years, children begin to prioritize the eyes for understanding the mental states of others, resulting in more sophisticated judgments of expression. This finding is consistent with study by Tonks, Williams, Frampton, Yates, and Slater (2007), who found that the ability to interpret complex mental state information from the eyes (Baron-Cohen, Wheelwright, & Joliffe, 1997) increases dramatically at age 11–12 years. The authors suggested that at this age, children begin to engage in increasingly complex social interactions, requiring greater sensitivity to others' emotional expressions. Thus, this age may represent an important stage in development for emotion expression comprehension, and our findings suggest that this may be related specifically to increased attention to the eyes.

Importantly, we found evidence for a specific bias to explore the left eye, relative to the right eye, that becomes significant in adulthood. Adults, but not children, spent more time exploring the left eye than the right eye. The adult left eye bias occurred very early on in viewing, with adults being highly likely to start their exploration by moving the window to the left eye, significantly more so than children. Thus, we suggest that sometime after age 12 years and before adulthood there is an additional change in attentional strategy that manifests as a bias to explore the left eye of faces during FER. Note that while there was a trend toward a left eye bias in 11- to 12-year-olds in terms of overall exploration time, the bias was not significant and did not emerge for the first ROI analysis, as it did for adults. Thus, a bias for the left eye may begin to emerge at 11–12 years, but it does not reach significance until a later age.

An attentional bias for the left half of faces in adulthood has been reported in eye tracking studies (Butler et al., 2005; Everdell, Marsh, Yurick, Munhall, & Paré, 2007; Hsiao & Cottrell, 2009; Mertens, Siegmund, & Grusser, 1993), in which participants make more fixations to the left visual field (LVF) than to the right visual field (RVF) of faces, and are more likely to direct initial saccades to the left side of faces than to the right side of faces (Leonards & Scott-Samuel, 2005). Indeed, Everdell et al. (2007) found a specific bias to look at the eye region in the LVF of faces, even for horizontally mirrored faces in which both halves of the face were identical, thus ruling out facial asymmetry as an explanation. This attentional bias for the LVF of faces is thought to reflect top-down control of attention exerted by lateralized (right-hemisphere) face processing centers (e.g., Kanwisher, McDermott, & Chun, 1997; Moscovitch et al., 1976; Nakamura et al., 1999). Others have found support for the possibility that a LVF bias may index perceptual expertise for both faces and objects (Guo et al., 2009; Hsiao & Cottrell, 2009). Our finding of a LVF preference using an alternative measure of attention, the MWT, may attest to the robustness of this attentional bias.

An attentional bias for the LVF-eye is also consistent with findings of a *perceptual* bias for the LVF of faces. This effect has been found using variety of tasks involving *chimeric faces* (faces made of two halves from different pictures, joined together)—including judgments of facial attractiveness, identity, age, gender, and emotion expression (e.g., Brady, Campbell, & Flaherty, 2005; Burt & Perrett, 1997; Christman & Hackworth, 1993). Most relevant to this study, judgments of expression are influenced more by the LVF than the RVF of the face (Chiang, Ballantyne, & Trauner, 2000). This perceptual bias for the LVF of faces is likely to be supported by the same mechanisms that drive attention to the left side of faces (Butler et al., 2005; Everdell et al., 2007).

Our finding that children do not show an LVF attentional bias is somewhat at odds with previous findings of an LVF perceptual bias in 10-year-old children (Chiang et al., 2000), and even children as young as 5 years (Levine & Levy, 1986). In addition, our results conflict with Guo et al.'s (2009) finding that both infants and adults are biased to look at the LVF of faces. However, it should be noted that the LVF perceptual bias for faces in very young children has been found to be weaker than that of older children and adults (Levine & Levy, 1986), and in some cases fails to reach statistical significance (Balas & Moulson, 2011; Chiang et al., 2000). An examination of Guo et al.'s data reveals that the LVF fixation bias in infants was (numerically) smaller than that of adults and extended to other species and even to inanimate objects, whereas in adults the LVF bias was specific to human faces. Taken together, the present data and previous findings suggest that the LVF bias may undergo significant developmental changes in childhood, including an increase in magnitude and specificity for faces.

Despite these interesting developmental effects for attention to eyes, it should be emphasized that overall, all age groups showed a strong bias to explore the mouth. Although somewhat unexpected given that previous eye tracking studies have revealed a bias to attend to the eyes (e.g., Hall et al., 2010; Pelphrey et al., 2002; Sullivan et al., 2007), our finding of an overall mouth bias may have several possible explanations. First, as the faces in this study depicted open-mouth expressions, participants may have quickly learned that the mouth was in fact highly informative. Indeed, participants were most likely to end their exploration on the mouth, that is, at the time of response (Figure 6a). However, clearly observers also found the eyes to be informative, as they were much more likely to initially move the mouse toward the eyes than to the mouth (Figure 5). Second, as discussed in the Introduction, it is possible that some eye tracking studies may overestimate the importance of eyes to FER, because (a) fixation and attention can be dissociated and (b) fixation data are often recorded for several seconds, possibly leading to the inclusion of fixation data that do not contribute to perceptual judgments of emotion. Finally, it is clear that during free viewing of full-resolution faces, such as in most eye tracking studies, gaze is guided by both bottom-up visual information and top-down control of attention. The eyes contain high-contrast information (Kobayashi & Koshima, 1997) and are likely to attract attention in part because they are visually salient (Kennedy & Adolphs, 2010), at least when the face is presented in isolation (Birmingham, Bischof, & Kingstone, 2009). In contrast, in the MWT paradigm exploration is largely under strategic control because there is a lack of high-resolution peripheral information to capture attention in a bottom-up manner. This may reduce the amount of attention that is normally afforded to the eyes. Because this represents a fundamental difference between eye tracking and the MWT, future investigations should examine the extent to which low-level visual information guides attention to the eyes of faces (Birmingham et al., 2009).

Consistent with predictions based on previous research (Calder et al., 2000; Smith et al., 2005; Tanaka et al., 2012), we found that exploration patterns depended on the emotion of the face. Participants spent more time examining the eyes (both left and right) of fearful and angry faces than of happy and disgusted faces, supporting the idea that eyes are informative for decoding fearful and angry expressions. In addition, observers were more likely to end their exploration on the eyes of angry and fearful faces relative to happy and disgusted faces, again suggesting that participants relied more on the eyes for judging these emotions. The mouths also showed a consistent pattern, with participants exploring the mouth of disgusted faces the most, and more frequently ending exploration on the mouths of disgusted and happy faces than fearful and angry faces. Thus, overall, our data suggest that observers seek out optimal perceptual information for decoding expressions.

Given these findings, it was not surprising to find that the distribution of attention between eyes and mouth was significantly correlated with accuracy and that this depended on the emotion of the face. Specifically, increased accuracy for anger and fear was associated with a bias for the eyes; conversely, increased accuracy for disgust was associated with a bias for the mouth. Accuracy for happy faces was not associated with a bias for eyes or mouth, likely because participants were highly successful at identifying this emotion and thus there was little variance in accuracy. Overall, our findings suggest that participants who were successful at FER adapted their distribution of attention between eyes and mouth to pick up optimal information for the emotion. This extends the research by Smith et al. (2005) and is consistent with findings from previous eye tracking research (e.g., Wong, Cronin-Golomb, & Neargarder, 2005).

Our findings show some similarities with previous research on developmental trajectories of accuracy for emotion recognition. Similar to Vicari et al. (2000), performance for happy faces was near ceiling even for 5- to 6-year-olds, and showed minimal improvements with age. In contrast, fearful and angry faces were most difficult for the youngest children in our study and improved gradually with age. Our RT data paralleled the accuracy data. Interestingly, these patterns were not evident in the baseline condition, in which we found developmental improvements in accuracy and RT, but no interaction with emotion type. We suggest that this reflects the utility of the moving window paradigm for examining age-related changes in face processing strategies. Unlike in the baseline condition, in which faces were in full view, in the moving window condition, children had to actively explore the faces to collect perceptual information. This likely amplified the different strategies of children at various developmental levels by making the task more challenging and necessitating the selection of some perceptual details at the cost of others.

Despite being consistent with previous research in some respects, our findings are at odds with studies reporting that young children perform particularly poorly at identifying disgust (e.g., Camras & Allison, 1985; Vicari et al., 2000). For instance, in contrast to Vicari et al. (2000), who found that 5- to 6-year-olds were very poor at labeling disgust and that accuracy rose sharply with age, we found that even young children were fairly accurate at identifying disgust ($M = 0.86$) and that that there was only a stepwise increase in accuracy for disgust at age 7–8 years with no further developmental improvements. Although it is difficult to know what may account for these discrepancies, we may offer two suggestions. First, task differences may account for the divergent findings. Although this study used a four-alternative forced choice labeling design, Vicari et al. asked children to freely label emotions. While the forced choice design has some limitations (discussed below), free labeling relies on being able to spontaneously generate a verbal label for an emotion, something that 5- to 6-year-olds may find particularly difficult. Likewise, Camras and Allison (1985) required children to interpret the meaning of a story, infer the appropriate emotion for the story, and finally to choose the correct label or image. Again, it may be that young children are able to identify disgusted expressions as such when given a set of responses to choose from but have difficulty generating the emotion label spontaneously, especially when this requires interpreting complex semantic information.

Second, stimulus differences between studies may be important for interpreting discrepant findings. Disgust may have been particularly easy to recognize in our study because we used open-mouth faces, with disgusted faces depicting a mouth that was highly unique (i.e., protruding tongue; Figure 1a). This may have been especially helpful for the young children, who explored the mouth for longer than older children and adults. In contrast, both Vicari et al. (2000) and Camras and Allison (1985) used photos depicting Ekman and Friesen's (1976) Facial Action Coding System, in which the facial action unit for disgust is a wrinkled nose. Thus, disgust may have been more difficult to recognize in these previous studies based on stimulus information alone. The issue of stimulus and task differences across studies is one that deserves systematic consideration in future research. It will be important to determine whether our present findings with the MWT generalize to closed-mouth faces or stimulus sets from other laboratories.

Forced choice paradigms, although advantageous in some respects, do have limitations in terms of interpreting results. An advantage of our design is that it restricted the possible emotion labels participants could choose from, thus reducing memory load and the need to generate a label from a large number of possibilities. However, it has been argued that constraining responses to a limited set without providing a "none of the above" option may produce artificial agreement among observers (Frank & Stennett, 2001; Russell, 1993). In addition, by constraining responses to only four emotions, we lost information about the potential confusions that participants may make among facial expressions.

For instance, previous research has found that fear is often confused for surprise (Palermo & Coltheart, 2004) especially in children (Gagnon, Gosselin, Hudon-ven der Buhs, Larocque, & Milliard, 2010). Without surprise as a response choice, the types of errors that participants made when viewing fearful faces do not necessarily reflect the common types of errors made in a more realistic setting. However, consistent with previous research using more response options (Gagnon et al., 2010), we found that 5- to 6-year-olds were more likely than older children to label disgusted faces as angry, suggesting that our paradigm picked up on at least one of the typical confusions made among emotions. Our study, like many others, has the problem that the conclusions are limited by the specific set of emotions included in the stimulus–response sets, and it will be important for future research to implement the MWT using different response paradigms to determine the robustness of the findings.

The MWT is not a replacement for eye tracking, and it has distinct limitations that constrain its use in developmental research. First, while we used window position to index the locus of attention, there is the possibility that some observers looked outside the window (e.g., when distracted), making window position an inaccurate measure of attention. However, we find this to be unlikely, as previous research has confirmed that observers commit the vast majority of their fixations to the center of the window in the MWT when used with complex scenes (Dalrymple et al., 2011). A possible solution would be the gaze-contingent paradigm, in which the window moves in conjunction with the participant's gaze. However, a difficulty with gaze-contingent paradigms is that the viewing experience may be uncomfortable, because the observer is forced to make saccades to empty space (Foulsham et al., 2011). Nonetheless, it will be important for future studies to validate the MWT for face stimuli by combining the MWT with eye tracking, to ensure that participants indeed look within the window boundaries.

Finally, the MWT also induces an analytic encoding strategy that is counter to the holistic encoding mechanism that has been implicated in expression perception (e.g., Calder & Jansen, 2005; but see Tanaka et al., 2012). Thus, it is not clear whether the attentional strategies observed in this study generalize to natural viewing conditions in which the face is perceived as whole. Despite this restriction, we argue that disrupting the holistic strategies may be necessary to gain a better understanding of attentional strategies underlying emotion recognition. Given that faces can be efficiently processed with very few fixations (e.g., Hsiao & Cottrell, 2008), measuring attention under natural viewing conditions may be inadequate for revealing how observers prioritize perceptual information for emotion recognition. That is, it may be necessary to slow down the process by forcing observers to actively explore the face to collect perceptual information relevant to the task. Furthermore, it is possible that a holistic representation may be constructed over the course of exploration (for a similar argument, see Caldara et al., 2010).

Despite these potential limitations, the MWT exhibits several advantages compared to eye tracking, including cost, participant comfort, and data collection and interpretation. The MWT is inexpensive, requiring only a computer and a mouse. It is nonintrusive and likely within the comfort zone of most children who have any experience with computers. Unlike eye tracking, the MWT does not require calibration, which can be both taxing for the participant and for the experimenter, who may have to recalibrate during the experiment or toss out trials if calibration drifts over time. Because visual information is restricted to the position of the window, by definition the researcher knows exactly what the participant is visually processing at any given moment. Finally, the MWT can be easily adapted to meet specific research goals. Indeed, we demonstrate one application of the MWT here, revealing a potentially important development shift in strategic allocation to faces during emotion recognition.

In sum, this study used a new approach, the *MWT*, to examine the attentional strategies involved in facial expression recognition. Our finding of a developmental increase in attention to eyes extends the literature on the development of FER. In particular, at age 11–12 years there begins to emerge a quantitative change in attention to the left eye, which matures into a significant left-eye bias by adulthood. Future studies should examine when the left-eye bias matures in adolescence and whether this enhanced attention to the eyes leads selectively to increased ability to read more complex states, especially those that rely on information afforded by the eye region (Baron-Cohen et al., 1997).

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*, 169–177.

Adolphs, R., & Birmingham, E. (2011). Neural substrates of social perception. In A. J. Calder, G. Rhodes, J. V. Haxby, & M. H. Johnson (Eds.), *Handbook of face* *perception* (pp. 571–589). Oxford, England: Oxford University Press.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*, 68–72.

Amso, D., Fitzgerald, M., Davidow, J., Gilhooly, T., & Tottenham, N. (2010). Visual exploration strategies and the development of infants' facial emotion discrimination. *Frontiers in Psychology, 1*, 1–7.

Bal, E., Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of Autism and Developmental Disorders, 40*, 358–370.

Balas, B., & Moulson, M. C. (2011). Developing a side bias for conspecific faces during childhood. *Developmental Psychology, 47*, 1472–1487.

Baron-Cohen, S., Wheelwright, S., & Joliffe, T. (1997). Is there a "language of the eyes?" Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition, 4*, 311–331.

Birmingham, E., Bischof, W., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*, 341–355.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research, 49*, 2992–3000.

Boyatzis, C. J., Chazan, E., & Ting, C. Z. (1993). Preschool children's decoding of facial emotions. *Journal of Genetic Psychology, 154*, 375–382.

Brady, N., Campbell, M., & Flaherty, M. (2005). Perceptual asymmetries are preserved in memory for highly familiar faces of self and friend. *Brain and Cognition, 58*, 334–342.

Buitelaar, J. K., van der Wees, M., Swaab-Barneveld, H., & van der Gaag, R. J. (1999). Verbal memory and Performance IQ predict theory of mind and emotion recognition ability in children with autistic spectrum disorders and in psychiatric control children. *Journal of Child Psychology and Psychiatry, 40*, 869–881.

Burt, D. M., & Perrett, D. I. (1997). Perceptual asymmetries in judgments of facial attractiveness, age, gender, speech and expression. *Neuropsychologia, 35*, 685–693.

Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia, 43*, 52–59.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "Spotlight" reveals universal information use for face recognition. *PLoS ONE, 5*, e708.

Calder, A. J., & Jansen, J. (2005). Configural coding of facial expressions: The impact of inversion and photographic negative. *Visual Cognition, 12*, 495–518.

Calder, A. J., Keane, J., Young, A. W., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 527–551.

Camras, L., & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *Journal of Nonverbal Behaviour, 9*, 84–94.

Chiang, C. H., Ballantyne, A. O., & Trauner, D. A. (2000). Development of perceptual asymmetry for free viewing of chimeric stimuli. *Brain and Cognition, 44*, 415–424.

Christman, S. D., & Hackworth, M. D. (1993). Equivalent perceptual asymmetries for free viewing of positive and negative emotional expressions in chimeric faces. *Neuropsychologia, 31*, 621–624.

Corden, B., Chilvers, R., & Skuse, D. (2008). Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. *Neuropsychologia, 46*, 137–147.

Dadds, M. R., Perry, Y., Hawes, D. J., Merz, S., Riddell, A. C., Haines, D. J., et al. (2006). Attention to the eyes and fear-recognition deficits in child psychopathy. *British Journal of Psychiatry, 189*, 280–281.

Dalrymple, K. A., Birmingham, E., Bischof, W., Barton, J. J. S., & Kingstone, A. (2011). Experiencing simultanagnosia through windowed viewing of complex social scenes. *Brain Research, 1367*, 265–277.

De Sonneville, L. M. J., Verschoor, C. A., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., & Vranken, M. (2002). Facial identity and facial emotions: Speed, accuracy, and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology, 24*, 200–213.

de Wit, T. C. J., Falck-Ytter, T., & von Hofsten, C. (2008). Young children with autism spectrum disorder look differently at positive versus negative emotional faces. *Research in Autism Spectrum Disorders, 2*, 651–659.

Donker, A., & Reitsma, P. (2007). Young children's ability to use a computer mouse. *Computers and Education, 48*, 602–617.

Eggum, N. D., Eisenberg, N., Kao, K., Spinrad, T. L., Bolnick, R., Hofer, C., et al. (2011). Emotion understanding, theory of mind, and prosocial orientation: Relations over time in early childhood. *Journal of Positive Psychology, 6*, 4–16.

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

Everdell, I. T., Marsh, H., Yurick, M. D., Munhall, K. G., & Paré, M. (2007). Gaze behaviour in audiovisual speech perception: Asymmetrical distribution of face-directed fixations. *Perception, 36*, 1535–1545.

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: Evidence from asymmetric gaze-contingent windows. *Attention, Perception & Psychophysics, 73*, 266–283.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology, 80*, 75–85.

Gagnon, M., Gosselin, P., Hudon-ven der Buhs, I., Larocque, K., & Milliard, K. (2010). Children's recognition and discrimination of fear and disgust facial expressions. *Journal of Nonverbal Behavior, 34*, 27–42.

Gosselin, P., & Larocque, C. (2000). Facial morphology and children's categorization of facial expressions of emotion: A comparison between Asian and Caucasian faces. *Journal of Genetic Psychology*, *161*, 346–358.

Guo, K., Meints, K., Hall, C., Hall, S., & Mills, D. (2009). Left gaze bias in humans, rhesus monkeys and domestic dogs. *Animal Cognition*, *12*, 409–418.

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science*, *198*, 853–855.

Hall, J., & Bernieri, F. (Eds.) (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Erlbaum.

Hall, J. K., Hutton, S. B., & Morgan, J. J. (2010). Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cognition and Emotion*, *24*, 629–637.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry*, *47*, 1098–1106.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*, 998–1006.

Hsiao, J. H., & Cottrell, G. (2009). Not all visual expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science*, *20*, 455–463.

Hunnius, S., de Wit, T. C. J., Vrins, S., & von Hofsten, C. (2011). Facing threat: Infants' and adults' visual scanning of faces with neutral, happy, sad, angry, and fearful emotional expressions. *Cognition & Emotion*, *25*, 193–205.

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*, 231–255.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *12*, 18–23.

Jansen, A. R., Blackwell, A. F., & Marriott, K. (2003). A tool for tracking visual attention: The restricted focus viewer. *Behaviour Research Methods, Instruments, & Computers*, *35*, 57–69.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.

Kennedy, D. P., & Adolphs, R. (2010). Impaired fixation to eyes following amygdala damage arises from abnormal bottom-up attention. *Neuropsychologia*, *48*, 3392–3398.

Klein, R. M., & Pontefract, A. (1994). Does oculomotor readiness mediate cognitive control of visual attention—revisited! In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 333–350). Cambridge, MA: MIT Press.

Kobayashi, H., & Koshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*, 767–768.

Leonards, U., & Scott-Samuel, N. E. (2005). Idiosyncratic initiation of saccadic face exploration in humans. *Vision Research*, *45*, 2677–2684.

Levine, S. C., & Levy, J. (1986). Perceptual asymmetry for chimeric faces across the life span. *Brain and Cognition*, *5*, 291–306.

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, *8*, 99–117.

McConkie, G. W., & Rayner, K. (1975). Span of effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578–586.

Merin, N., Young, G. S., Ozonoff, S., & Rogers, S. J. (2007). Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants. *Journal of Autism and Developmental Disorders*, *37*, 108–121.

Mertens, I., Siegmund, H., & Grusser, O. J. (1993). Gaze motor asymmetries in the perception of faces during a memory task. *Neuropsychologia*, *31*, 989–998.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology*, *86*, 67–84.

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2008). The development of dynamic facial expression recognition at different intensities in 4- to 18-year-olds. *Social Development*, *19*, 71–92.

Moscovitch, M., Scullion, D., & Christie, D. (1976). Early versus late stages of processing and their relation to functional hemispheric asymmetries in face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 401.

Nakamura, K., Kawashima, R., Ito, K., Sugiura, M., Kato, T., Nakamura, A., et al. (1999). Activation of the right inferior frontal cortex during assessment of facial emotion. *Journal of Neurophysiology*, *82*, 1610–1614.

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., et al. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society, B*, *277*, 2935–2943.

Neumann, D., Spezio, M. L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: Abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social Cognitive and Affective Neuroscience*, *1*, 194–202.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods*, *36*, 634–638.

Pelphrey, K. A., Sasson, N. J., Reznick, S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*, 249–261.

Rump, K. M., Giovannelli, J. L., Minshew, N. J., & Strauss, M. S. (2009). The development of emotion recognition in individuals with autism. *Child Development, 80*, 1434–1447.

Russell, J. A. (1993). Forced-choice response format in the study of facial expressions. *Motivation and Emotion, 17*, 41–51.

Smilek, D., Frischen, A., Reynolds, M. G., Gerritsen, C., & Eastwood, J. D. (2007). What influences visual search efficiency? Disentangling contributions of preattentive and postattentive processes. *Perception & Psychophysics, 69*, 1105–1117.

Smith, M. L., Cottrell, C. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*, 184–189.

Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signalling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology, 21*, 195–200.

Steele, H., Steele, M., & Croft, C. (2008). Early attachment predicts emotion recognition at 6 and 11 years old. *Attachment & Human Development, 10*, 379–393.

Sullivan, S., Ruffman, T., & Hutton, S. B. (2007). Age differences in emotion recognition skills and the visual scanning of emotion faces. *The Journals of Gerontology, 62*, P53–P60.

Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: The holistic and analytic perception of facial expressions. *Cognition & Emotion 26*, 961–977.

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science, 10*, 547–558.

Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9–10 year-olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury, 21*, 623–629.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*, 242–249.

Vicari, S., Reilly, J. S., Pasqualetti, P., Vizzotto, A., & Caltagirone, C. (2000). Recognition of facial expressions of emotions in school-age children: The intersection of perceptual and semantic categories. *Acta Paediatrica, 89*, 836–845.

Wallace, S., Coleman, M., & Bailey, A. (2008). An investigation of basic facial expression recognition in autism spectrum disorders. *Cognition & Emotion, 22*, 1353–1380.

Wong, B., Cronin-Golomb, A., & Neargarder, S. (2005). Patterns of visual scanning as predictors of emotion identification in normal aging. *Neuropsychology, 19*, 739–749.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1425–1441

Daily Intragroup Contact in Diverse Settings: Implications for Asian Adolescents' Ethnic Identity

Tiffany Yip and Sara Douglass
Fordham University

J. Nicole Shelton
Princeton University

This study examined the daily-level association between contact with same-ethnic others and ethnic private regard among 132 Asian adolescents (mean age = 14 years) attending four high schools ranging in ethnic composition diversity. The data suggest a positive daily-level association between contact with same-ethnic others and ethnic private regard for adolescents, who were highly identified with their ethnic group and who attended predominantly White or ethnically heterogeneous schools. In addition, using time lag analyses, contact with same-ethnic others yesterday was positively related to ethnic private regard today, but ethnic private regard yesterday was unrelated to contact with same-ethnic others today, suggesting that adolescents' identity is responsive to their environments. The implications of these findings for the development of ethnic identity are discussed.

Adolescents encounter a variety of people on a daily basis; some of these people are members of the adolescent's ethnic group. These daily encounters are likely to have implications for the development of adolescents' identity in general, and for ethnic minority adolescents, these daily encounters are likely to shape their ethnic identity more specifically. Indeed, research shows that same-ethnic peers serve as important socializing agents by providing information about what it means to be a member of one's ethnic group (Hamm, 2000; Lee, Noh, Yoo, & Doh, 2007; Tatum, 2004; Umana-Taylor, 2004; Yip, Seaton, & Sellers, 2010). However, little research has been conducted on how the interaction between personal and structural characteristics shapes the relation between contact with same-ethnic others and ethnic identity in adolescents' daily lives.

This study examines the daily-level relation between intragroup contact and Asian adolescents' ethnic private regard—the feelings associated with being a member of one's ethnic group (Sellers, Rowley, Chavous, Shelton, & Smith, 1997). In service of this goal, we acknowledge three distinct components of this daily-level relation. First, we pay particular attention to the directionality of the daily-level association between intragroup contact and ethnic private regard. Namely, does the amount of intragroup contact yesterday predict ethnic private regard today, suggesting that intragroup contact shapes subsequent identity development? Or, does ethnic private regard yesterday predict the amount of intragroup contact today, suggesting that adolescents have agency (i.e., an adolescent's sense of competence, instrumentality, and autonomy to make decisions that influence his or her own development; Bandura, 2006) in shaping their social contexts by seeking contact with similar others when they are feeling positive about their ethnic identity? Second, we examine the extent to which Asian adolescents' ethnic identity centrality—the level of importance of one's ethnic group membership—influences the daily-level association between intragroup contact and ethnic private regard. Third, we examine the extent to which structural factors—in this case, the ethnic composition of adolescents' high school—moderate these relations. Consistent with scholarship on Asians experience in the United States, we use the label *ethnic identity* (as opposed to *racial identity*) in our work because ethnicity has been observed to be more salient to Asians in the United States (Espiritu, 1992; Uba, 1994). However,

---

This research was supported by a grant awarded to the first and third authors from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (1 R01 HD055436).

Correspondence concerning this article should be addressed to Tiffany Yip, Department of Psychology, Fordham University, Bronx, NY 10458. Electronic mail may be sent to tyip@fordham.edu.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0022
DOI: 10.1111/cdev.12038

we use the terms *ethnic* and *racial identity* interchangeably or in combination (*ethnic or racial*) when referring to previous research to be consistent with the way previous researchers have used the terms in their work.

We focus on the identity experiences of Asian adolescents for several reasons. First, although Asians represent one of the smaller minority groups in the United States (5% compared to 16% Latino and 12% Black), they are also one of the fastest growing groups (U.S. Census Bureau, 2010). Second, although Asians are a unique group with cultural values and stereotypes that set them apart from other ethnic groups, there is also a great deal of diversity within the broader group (Espiritu, 1992). As such, the importance of examining the effect of intragroup contact on ethnic identity is particularly relevant because contact with other Asians can either emphasize commonality ("We are all Asian") or difference ("You are Korean and I am Indonesian"). Finally, research on discrimination among Asian American adolescents finds that negative intragroup exchanges are more common compared to other ethnic groups (Fisher, Wallace, & Fenton, 2000; Rivas-Drake, Hughes, & Way, 2008), underscoring the importance of examining how intragroup contact influences Asian adolescents' identity development.

We draw upon ecological theory as a developmental framework for understanding how personal and structural characteristics interact to influence daily processes associated with ethnic identity (Bronfenbrenner, 1999). This framework allows for the developing person to affect the context as well as a reciprocal relation in which the context influences the person. We adopt this framework to address the daily association between intragroup contact and adolescents' feelings about their ethnic group, taking into account both personal and structural characteristics.

In focusing on social identity processes, we also draw upon self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987), which states that all individuals belong to multiple social categories (e.g., female gender, Asian, etc.) and that the interplay between personal characteristics and one's immediate structural context determines which of those categories is salient at a specific point in time. "Categorization-in-context" (Turner et al., 1987) considers social identities to be fluid, dynamic, and dependent upon one's immediate context. The composition of identities in a context and the social interactions that occur within that context create a *comparative context*, which determines the salience of one's social identities. In this study, we explore how racial composition of adolescents' daily context and the contact they have with ethnic in-group members shape their ethnic private regard.

Private Regard: Feelings About One's Ethnic Group

Building off of self-categorization theory (Turner et al., 1987), research on ethnic identity finds that the importance of ethnicity and how individuals feel about being a member of their ethnic group are stable and dynamic (Sanchez & Garcia, 2009; Sanchez, Shih, & Garcia, 2009; Sellers, Smith, Shelton, Rowley, & Chavous, 1998; Yip, 2005). Until recently, there has been little research on how one's ethnic identity varies across contexts (Forehand, Deshpande, & Reed, 2002; Hutnik & Sapru, 1996; Shih, Pittinsky, & Ambady, 1999; Yip, 2005). Moreover, the existing work includes adult and young adult samples, not adolescents, despite adolescence being a primary period of identity construction (Erikson, 1968; Phinney, 1993).

People generally evaluate in-group members more positively than out-group members (see Brewer, 2007, for a review). They are more likely to feel that they experience shared fate with (Gurin & Townsend, 1986) and are understood by (Seder & Oishi, 2009) in-group members. Given this attachment to and interdependence with in-group members, encounters with similar others may make people's positive regard for the in-group salient because they know that they are being evaluated positively and are understood. However, encounters with similar others are also likely to highlight the diversity within one's ethnic group, causing people to wrestle with how they are being perceived by other in-group members (Branscombe, Spears, Ellemers, & Doosje, 2002). For example, African Americans and Asian Americans, who are perceived as "acting White" and do not conform to their group norms, are marginalized by in-group peers (Neal-Barnett, Stadulis, Singer, Murray, & Demmings, 2010). The extent to which people feel they are accepted or rejected by in-group members may influence their feelings of private regard. For example, foreign exchange students who feel rejected by their host country, compared to those who feel accepted by the country, are more likely to disidentify with the country (i.e., want to leave; Matschke & Sassenberg, 2010). Next we discuss how the variation in the type of intragroup contact—active versus passive—might influence ethnic private regard.

Intragroup Contact and Self-Perceptions

People prefer to be surrounded by and interact with similar others, perhaps because they feel more comfortable and less distressed when in contact with people like themselves (McPherson, Smith-Lovin, & Cook, 2001). As suggested by self-categorization theory (Turner et al., 1987), this may be especially true for people who are in settings where similar others are not readily available. For individuals who find themselves in the numerical minority in a setting (e.g., ethnic minority in a predominantly White school), the presence of out-group members may increase in-group affiliation and well-being. Indeed, for stigmatized and ethnic minorities who often find themselves in the numerical minority in situations, being in the presence of similar others is associated with positive psychological well-being and self-perceptions (Frable, Platt, & Hoey, 1998; Sanchez & Garcia, 2009). For example, in an experience sampling study of college students, participants who were members of concealable socially stigmatized groups (e.g., gay or lesbian) reported feeling better about themselves, less anxious, and less depressed when they were with similar others (Frable et al., 1998). Likewise, an experience sampling study including multiethnic adults found that individuals reported feeling more valued when other minorities were present (Sanchez & Garcia, 2009). The relations found in both studies may occur because similar others typically have more positive perceptions of one's group and are able to provide support to help negotiate the challenges of one's group membership (Frable et al., 1998).

Although the aforementioned studies are informative, they do not offer insight into how encounters with similar others shape adolescents' own evaluation about being a member of their stigmatized group. In addition, across these studies, intragroup contact has been conceptualized in multiple ways that vary in degree of direct contact, leaving unanswered the question of which type of contact has a stronger impact on ethnic minorities' self-perceptions. In the present research, we examine how passive (being surrounded by) and active (interacting with) intragroup contact are related to adolescents' evaluations about being a member of their ethnic group on a daily basis, including examining the directionality of the relationships.

Interacting With Versus Being Surrounded by Same-Ethnic Others

Contact with same-ethnic others can take many forms, and the operationalization of intragroup contact has varied across previous studies. Some researchers construe intragroup contact as social interactions that suggest people are engaged in an activity together (e.g., socializing). This type of contact is usually measured with items such as, "Who are you with right now?" (Frable et al., 1998). Other researchers construe intragroup contact by including anyone in a target's general surrounding, regardless of whether the people are engaged in an activity with one another. This type of contact is usually measured with items such as, "How many people of Asian descent are around you?" (Sanchez & Garcia, 2009). Although each type of intragroup contact is related to psychological well-being, it is not clear if one has a stronger effect than the other because both have not been included in the same study as independent constructs. Simply being surrounded by similar others (e.g., having Asian students in class) does not imply the same level of involvement as interacting with similar others (e.g., talking with Asian students in class). On one hand, intragroup contact may require a certain level of involvement with same-ethnic others to influence ethnic identity. On the other hand, simply being surrounded by same-ethnic others may be comforting for individuals who generally find themselves in contexts where they are in the numerical minority, and thus, active contact may not be necessary to influence ethnic identity. In the present research, we include two conceptually distinct, yet likely related, measures of intragroup contact: (a) active contact—engaging in an activity during an interaction with other individuals (e.g., talking to someone on the subway) and (b) passive contact—merely being surrounded by other individuals, but not engaging in an activity together (e.g., riding the subway with a group of people). By including both forms of intragroup contact, we approach a more nuanced understanding of how contextual factors shape adolescents' ethnic identity.

Directionality: Does Intragroup Contact Yesterday Predict Private Regard Today or Does Private Regard Yesterday Predict Intragroup Contact Today?

The psychological literature has conceptualized intragroup contact as predicting self-appraisals; contact with other in-group members predicts subsequent feelings about oneself (Tatum, 1997), and existing research supports this notion (Phinney, Romero, Nava, & Huang, 2001; Tatum, 1997). However, in the absence of within-person repeated measures, it is not clear if the relation is reciprocal. If adolescents have agency over their identity development, those who felt especially positive about their ethnic identity yesterday might seek to affiliate with more in-group individuals today. Moreover, there may be a bidirectional relation where contact with Asians yesterday predicts higher private regard today *and* higher private regard yesterday predicts contact with Asians today. We address directionality in this study.

Moderators of the Relation Between Intragroup Contact and Ethnic Private Regard

Both ecological and self-categorization theories (Turner et al., 1987), suggest that the relation between daily intragroup contact and ethnic private regard is likely to be moderated by personal and contextual factors. We examine ethnic identity centrality as a possible individual-difference moderator, and ethnic composition of the context as a possible contextual moderator, of the relation between intragroup contact and ethnic private regard.

Personal factors: Ethnic identity centrality. People are more aware of environmental cues that are relevant (vs. irrelevant) to important social identities (Turner et al., 1987). Consistent with this notion, ethnic identity centrality serves as a lens through which people experience the world (Sellers et al., 1998; Shelton & Sellers, 2000). Indeed, people who are highly identified with their ethnic group interpret ethnically ambiguous situations as ethnically motivated (Shelton & Sellers, 2000). Using experience sampling methods, Yip (2005) found that ethnic identity centrality determined how Asian American college students experienced their proximal environments, such that those who were higher on centrality reported higher ethnic identity salience on days in which they were surrounded by more Asian individuals. Building upon this study, we predict that the relation between intragroup contact and ethnic private regard may differ depending on how important being Asian is to an adolescents' identity. We contend that for Asian adolescents who report that being Asian is central to their identity, having contact with other Asians will be associated with more positive feelings about being Asian. However, for Asian adolescents who report that being Asian is not central to their identity, contact with other Asians will be irrelevant for one's feelings about being Asian. If ethnic identity centrality is a lens through which the world is viewed, then adolescents who choose not to make race important to their identity should report that contact with other in-group members has no bearing on how they feel about being a group member. The effect of ethnic centrality on the daily associations between intragroup contact and ethnic private regard is likely to be further influenced by characteristics of the larger context.

Structural factors: Ethnic composition of context. Researchers interested in examining adolescent development in natural settings have gravitated toward perspectives that provide contextualized approaches to studying development (Barker, 1968; Gallimore & Goldenberg, 1993; O'Donnell, Tharp, & Wilson, 1993; Super & Harkness, 1986; Whiting & Whiting, 1975). The importance of context has been especially highlighted in the area of social identity research (e.g., Ashmore, Deaux, & McLaughlin-Volpe, 2004; Branscombe, Deaux, & Lerner, 1985; Byrd & Chavous, 2009; Chavous, Harris, Rivas, Helaire, & Green, 2004; Ethier & Deaux, 1994, 2001). Unfortunately, although self-categorization theory (Turner et al., 1987) underscores the importance of the immediate context, previous research on intragroup contact and self-perceptions has not taken into consideration the different contexts in which contact occurs. Instead, the focus has been on people who are in contexts in which their group is normally in the numerical minority (e.g., biracial student at predominantly White college). Sometimes ethnic minorities are not in the numerical minority, but rather are in the numerical majority, on a normal basis. For example, a Chinese adolescent who lives in a predominantly Asian neighborhood, attends a predominantly Asian high school, and participates in activities with predominantly Asian peers, may find herself constantly surrounded by other Asians. Does being surrounded by Asians relate to self-perceptions for this young girl? To our knowledge, research is lacking on how intragroup contact influences ethnic self-perceptions when ethnic minorities are in the numerical majority or in ethnically heterogeneous environments.

We are interested in the ethnic composition of adolescents' schools. Adolescents attend schools where their ethnic group may be in the numerical minority or the numerical majority. The ethnic composition of schools plays a role in the relation between contact with same-ethnic peers and ethnic identity. For example, African American college students who attended predominantly White high schools reported feeling isolated from their White peers in high school, and searched for contact with African Americans outside of school to develop a racial identity (Tatum, 2004). Examining annual changes in racial identity, Yip, Seaton, and Sellers (2010) found that contact with same-race peers and racial identity stability varied as a function of the students' school racial composition. For African American adolescents attending racially diverse schools (i.e., relatively equal numbers of minority groups and Whites), less contact with African American peers was related to racial identity instability over a 3-year period. For African Americans attending predominantly White schools, however, more contact with same-race peers was related to identity stability over time. Thus, being in the racial or ethnic minority or majority in school has implications for ethnic identity.

Since Asians comprise under 5% of the U.S. population (U.S. census bureau, 2010), many Asian adolescents attend schools where their ethnic group is in the numerical minority. In these contexts, people are apt to experience negative psychological outcomes (Niemann & Dovidio, 1998; Pollak & Niemann, 1998; Saenz, 1994) because being in the numerical minority is associated with feelings of social isolation (Martin & Hetrick, 1988), which in turn are related to feelings about one's group membership (Frable et al., 1998). When in minority contexts, interacting with or merely being surrounded by other Asians may reduce the negative affect associated with having one's group be in the numerical minority because adolescents will be reminded that others (in-group members) have positive views of their group and feel like they are understood. Some, although fewer, Asian adolescents attend schools in which their ethnic group is in the numerical majority. In these contexts, intragroup contact and ethnic private regard may be unrelated for adolescents because their group is generally perceived in a positive manner by the majority group (i.e., in-group members). Thus, daily intragroup contact is unnecessary as a buffer from the negative psychological outcomes associated with being in the numerical minority. Alternatively, attending a school in which Asians are in the majority may be negatively associated with ethnic private regard because Asian students may make distinctions among themselves (e.g., generational status) that could create divisiveness and problems (Lee, 1996). Social comparison theory (Festinger, 1954) suggests that individuals compare themselves to similar others in to evaluate themselves. In majority schools, "similarity" is likely defined by more nuanced characteristics (e.g., native language) that ironically lead people to create in-group distinctions, resulting in lower private regard among some individuals. Indeed, Turner (1985, p. 257) proposed that "personal self becomes salient where comparisons are restricted to in-group members." In an ethnographic study in a predominantly Asian high school, Lee (1996) found that Asian students derogatorily referred to recent immigrants as "fresh off the boat" to create a distinction from peers who were more "American."

This Study

Using ecological theory as a developmental framework and self-categorization theory (Turner et al., 1987) to focus on identity processes, we examine how personal and structural factors interact to influence the daily-level relation between intragroup contact and ethnic private regard for Asian adolescents. We focus on how the type of intragroup contact—being surrounded by versus interacting with same-ethnic others—may influence these relations, and explore the directionality of the relations. We predict that for Asian adolescents who report that their ethnicity is important to their overall identity, the more intragroup contact they experience the more positive their ethnic private regard. Moreover, we predict that in schools where Asian adolescents are in the numerical minority, being surrounded by and interacting with same-ethnic peers will be related to more positive ethnic private regard because of the sense of comfort and belonging that similar others can provide. Furthermore, we expect these associations to be especially strong for adolescents who report high levels of ethnic identity centrality. We predict a weaker association in schools where Asians are in the numerical majority. Finally, consistent with research (Hartup, 2005; Levin, van Laar, & Sidanius, 2003), we predict that intragroup contact will influence subsequent ethnic private regard. However, as this is a new area of inquiry, we do not have hypotheses about the bidirectionality of this relation.

Method

Participants

Asian adolescents participated in a study on identity development ($N = 132$; $M_{\text{age}} = 14.12$, $SD = 0.72$; 92 female participants). In an open-ended question about ethnicity, 59 participants reported they were Chinese (e.g., Cantonese, Chinese), 45 participants reported they were Asian (e.g., Asian, Asian American), and the remaining participants ($n = 28$) reported other Asian ethnicities (e.g., Bengali, Filipino). Approximately three fourths of the sample reported being born in the United States ($n = 97$), with most adolescents ($n = 94$) reporting at least 1 foreign-born parent and the remaining 3 adolescents reporting U.S.-born grandparents. Of participants who reported being born in other countries ($n = 35$), the majority reported being born in China ($n = 22$), Bangladesh ($n = 3$), and the Philippines ($n = 2$), and age of immigration ranged from 6 months to 14 years old ($M = 6.36$, $SD = 4.54$). A majority of participants reported not knowing the highest level of education completed by their parents ($n = 70$), but the most common response of those who knew was that their parents completed high school ($n = 20$).

Procedure

We obtained data on the ethnic composition of the public high schools from the New York Department of Education (DOE). On the basis of these data, we selected four similar-sized and academically comparable schools to represent a predominantly Asian school ($n = 1$), a predominantly White school ($n = 1$), and ethnically heterogeneous schools ($n = 2$; Table 1). *Predominantly* is defined by a single group representing at least 40% of the school's student population. In the ethnically heterogeneous school, no ethnic group represented more than 40% of the school's population. Once the school administrators agreed to participate in the study, we sent parental consent and youth assent letters home to all ninth graders. We included only students with completed consent and assent forms. In this article, we focus solely on Asian adolescents because we are interested in examining ethnic minorities when they are in a context in which they are in the numerical majority compared to the numerical minority. We are unable to do that with the other ethnic minority students in this study.

At each school, we administered surveys to participants during an orientation session in groups ranging from 10 to 30 students. These surveys included demographic questions along with measures of ethnic identity. After participants completed the surveys, we gave them a cellular phone to access a web-based survey every night for 7 days before going to bed. All functions on the cellular phone were blocked with the exception of accessing the study's web survey. We provided participants with detailed instructions on how to use the phones and complete the surveys. We tracked compliance online each night. At the end of the week, participants were required to return the cellular phone to receive $50 for their participation. On average, participants completed 6.90 (range = 5 to 7) surveys over the course of the week.

Measures

Stable Ethnic Identity

We adapted the private regard and centrality subscales of the Multidimensional Inventory of Black Identity (MIBI; Sellers et al., 1997) for use with a multiethnic sample. The Private Regard subscale has seven items that measure one's feelings about being a member of his or her racial or ethnic group (e.g., "I feel good about people from my racial or ethnic group"; $M = 5.36$, $SD = 0.90$, $\alpha = .79$). The Centrality subscale has eight items that measure the importance of race or ethnicity to one's overall self-concept (e.g., "In general, my race or ethnicity is an important part of my self-image"; $M = 4.53$, $SD = 0.87$, $\alpha = .69$). Participants completed the subscales using response scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Since stable ethnic identity influences daily-level ethnic identity (i.e., how individuals feel about their ethnic group in general affects how they feel about their ethnic group on a specific day; Yip & Fuligni, 2002), it was important to control for stable ethnic private regard in all analyses.

School-Reported Ethnic Composition Index

On the basis of data available on the New York City DOE website about the percentage of White, Black, American Indian, Asian, Hispanic, and Other

Table 1
School Ethnic Compositions

School	Total sample %	Asian in school %	Black in school %	Latino in school %	White in school %
Predominantly Asian	33	58	7	11	24
Predominantly White	38	28	9	14	47
Heterogeneous	24	28	13	24	34
Heterogeneous	5	16	22	28	33

students at each school, we created an objective, school-reported diversity index. Employing methods described in Juvonen, Nishina, and Graham (2006), we computed the ethnic diversity of each student's school based on the percentage of students in each ethnic group. Using the number of different ethnic groups (g) and the proportion of individuals (p) who are members of each group (i), the index (D_C) provides an estimate of the relative probability that two randomly selected students are from different ethnic groups:

$$D_C = 1 - \sum_{i-1}^{g} p_i^2$$

Higher scores (range = 0 to 1) indicate having more Black and White, but fewer Asian, students (Asians in numerical minority), whereas lower scores indicate the presence of more Asian, but fewer Black and White, students (Asians in numerical majority; $M = 0.65$, $SD = 0.06$).

Daily Ethnic Identity

To assess daily ethnic private regard, the MIBI–S Private Regard scale was adapted for how participants felt that day (Martin, Wout, Nguyen, Sellers, & Gonzalez, 2005). Participants completed two items (e.g., "I was happy that I am my race or ethnicity"; $M = 3.72$, $SD = 1.41$, $\alpha = .87$). To assess daily centrality, participants completed four Centrality items (e.g., "In general, my race or ethnicity was an important part of my self-image"; $M = 3.38$, $SD = 1.25$, $\alpha = .86$), with response scales ranging from 1 (*not at all*) to 7 (*extremely*).

Daily Intragroup Contact

During the orientation session, we explained to participants, using the people attending the session as a demonstration, how to distinguish between the two types of intragroup contact constructs. We made sure that participants understood the distinction before moving forward with completing the measures. To assess the amount of active contact with same-ethnic others, participants responded to the following: "Think about all the people who you *interacted with* today. How many were the same race or ethnicity as you?" ($M = 2.34$, $SD = 1.12$). An example of active contact included "someone you talked to today." To assess the amount of passive contact with same-ethnic others, participants responded to the following: "Think about all the people who were *surrounding you* today. How many were the same race or ethnicity as you?" ($M = 2.22$, $SD = 1.12$). An example of passive contact included "someone who was in the same area as you (e.g., cafeteria) but with whom you were not talking." Participants made their responses using the following scale: 1 = *none: 0%*, 2 = *about 25%*, 3 = *about 50%*, 4 = *about 75%*, 5 = *all: 100%*.

Control Variables

Given their theoretical and statistical relations with our primary variables, we controlled for the following variables in our analyses.

Day of study. A change in daily ethnic private regard over the course of the study may represent an artifact of the diary method (i.e., explicitly thinking about these feelings on a regular basis may influence people's responses). Therefore, we controlled for this in our analyses. Also, because daily ethnic private regard was correlated with day of the study ($r = -.71, p < .10$), we adjusted for the variable along with type of day (i.e., weekday vs. weekend).

Perceived ethnic composition of contexts. Perceived school ethnic composition was included in the models to account for how participants perceived their school contexts so that the effects of school ethnic composition could be isolated more accurately. We asked adolescents "What percentage of people in your school are: African American or Black, Asian American or Asian, Hispanic or Latino, White, Other, People I cannot identify." Although the focus in this article is on adolescents' school contexts, we did not restrict the measurement of intragroup contact to school interactions. It was important to account for other proximal settings in which adolescents are likely to interact with same-ethnic others; therefore, perceived neighborhood ethnic composition was included by asking "What percentage of people in your neighborhood are…."

Demographic control variables. We included gender and immigration status in the initial models; however, they were removed because they did not affect the overall models.

Results

To examine the correlations between person-level and daily-level variables, we computed the mean of the daily variables across the 7 days of the study (Table 2). Consistent with existing research (e.g., Sellers et al., 1997), there were significant relations between stable and daily centrality and private

Table 2
Bivariate Correlations Among Study Variables

	1	2	3	4	5	6	7	8	9
1. Stable centrality	—	.53***	−.14	−.04**	−.08	.54***	.40***	.19*	.15
2. Stable private regard		—	−.12	−.17$^+$	−.20*	.46***	.67***	.22*	.12
3. Department of Education (DOE)–reported school ethnic comp.			—	.14	−.03	−.15	−.05	−.13	−.16
4. Perceptions of school ethnic comp.				—	.32**	−.25**	−.22*	−.07	−.09
5. Perceptions of neighborhood ethnic comp					—	−.11	−.13	−.26**	−.27**
6. Daily centrality						—	.74***	.26**	.27**
7. Daily private regard							—	.31**	.23*
8. Daily contact—interacted with								—	.85**
9. Daily contact—surrounded by									—

Note. School comp. = school ethnic composition; Neighborhood comp. = neighborhood ethnic composition.
$^+p < .10$. $^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

regard. The passive and active intragroup contact measures were highly related ($r = .85$, $p < .01$). No mean differences were found by gender, ethnic self-categorization, immigration status, or school setting for any of the stable or daily ethnic identity scores.

Since school context (predominantly Asian, predominantly White, or ethnically heterogeneous) could influence the opportunity for intragroup contact, we examined possible mean differences in intragroup contact. There were no differences in passive intragroup contact. However, there was a difference in active intragroup contact, $F(2, 118) = 4.19$, $p < .05$, such that adolescents in the predominantly Asian school reported higher mean levels of active intragroup contact ($M = 2.61$) than adolescents in the heterogeneous schools ($M = 1.59$).

Data Analyses Overview

The nested structure of the data required hierarchical linear models (HLM; Bryk & Raudenbush, 1992). HLM is appropriate for nested data as it allows for simultaneous analyses of data at more than one level. For the current research, it allows investigation into the moderating effects of both personal-level (e.g., identity) and structural-level (e.g., school composition) variation on daily experiences. In this study, we employ a three-level model as multiple daily diaries are nested within individuals, who are in turn nested within schools. Level 1 represents daily-level experiences, Level 2 represents adolescents' stable characteristics, and Level 3 represents school characteristics. Notably, although self-reported school (and neighborhood) diversity is based on school-level phenomenon, it represents each adolescent's perception of his or her school and, thus, is treated as a person-level variable (Level 2).

Concurrent-Day Models

Daily Intragroup Contact and Private Regard

First, we explored the daily-level relation (Level 1 only) between intragroup contact (interacted with and surrounded by same ethnic others) and private regard without considering personal- and structural-level (Level 2 and Level 3, respectively) moderators (Figure 1). We included daily ethnic centrality as a control variable to account for systematic fluctuations that may occur with private regard. For the Level 1 model, P0 represents the intercept, and E represents the random component of private regard not accounted for in the model.

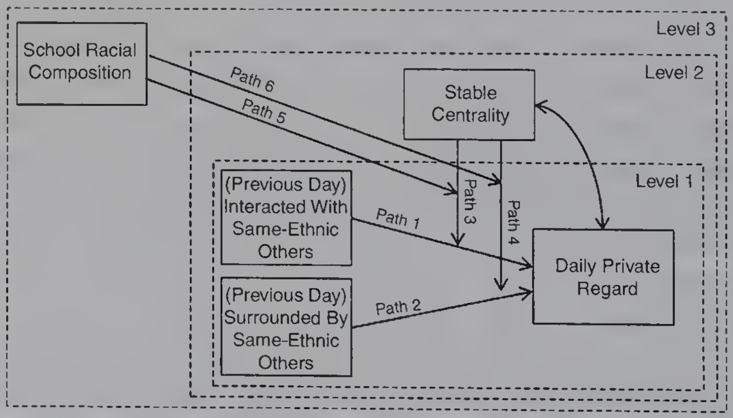

Level 3
School Racial Composition
Level 2
Path 6
Path 5
Stable Centrality
Level 1
(Previous Day) Interacted With Same-Ethnic Others
Path 3
Path 4
Path 1
Daily Private Regard
(Previous Day) Surrounded By Same-Ethnic Others
Path 2

Figure 1. Schematic for multilevel model investigating relation of same-day (and previous-day) interacting with and being surrounded by same-ethnic others with daily private regard. Parallel but independent models were run for same-day and previous-day analyses.

Level 1 model:

$$\begin{aligned}&\text{DAILY PRIVATE REGARD} =\\&\quad \text{P0} + \text{P1} * (\text{DIARY}\\&\quad \text{DAY}) + \text{P2} * (\text{WEEKDAY/WEEKEND}) + \text{P3} *\\&\quad (\text{DAILY CENTRALITY}) + \text{P4} * (\text{INTERACTED}) +\\&\quad \text{P5} * (\text{SURROUNDED}) + \text{E}\end{aligned}$$

Neither intragroup contact variable—interacting with (P4; Figure 1, Path 1) and being surrounded by (P5; Figure 1, Path 2) same-ethnic people—were associated with daily private regard.

Stable Centrality as a Moderator

Next, we included stable ethnic centrality at Level 2 to investigate influences on the daily-level association between intragroup contact and private regard (Figure 1, Paths 3 and 4). We also included stable private regard at Level 2 to isolate the daily fluctuations in private regard from stable evaluation of one's ethnic identity. We included all control variables as predictors of the intercept (P0) and the association between intragroup contact and daily private regard (P4, P5). Error terms are indicated by R, and represent the random components not accounted for in the model.

Level 2 Model:

$$\begin{aligned}\text{P0} &= \text{B00} + \text{B01} * (\text{PRIVATE REGARD}) + \text{B02} *\\&\quad (\text{CENTRALITY}) + \text{B03} * (\text{PERCEPTIONS OF}\\&\quad \text{SCHOOL ETHNIC COMPOSITION}) + \text{B04} *\\&\quad (\text{PERCEPTIONS OF NEIGHBORHOOD}\\&\quad \text{ETHNIC COMPOSITION}) + \text{R0}\\\text{P1} &= \text{B10}\\\text{P2} &= \text{B20}\\\text{P3} &= \text{B30} + \text{R3}\\\text{P4} &= \text{B40} + \text{B41} * (\text{PRIVATE REGARD}) + \text{B42} *\\&\quad (\text{CENTRALITY}) + \text{B43} * (\text{PERCEPTIONS OF}\\&\quad \text{SCHOOL ETHNIC COMPOSITION}) +\\&\quad \text{B44} * (\text{PERCEPTIONS OF NEIGHBORHOOD}\\&\quad \text{ETHNIC COMPOSITION}) + \text{R4}\\\text{P5} &= \text{B50} + \text{B51} * (\text{PRIVATE REGARD}) + \text{B52} *\\&\quad (\text{CENTRALITY}) + \text{B53} * (\text{PERCEPTIONS}\\&\quad \text{OF SCHOOL ETHNIC COMPOSITION}) +\\&\quad \text{B54} * (\text{PERCEPTIONS OF NEIGHBORHOOD}\\&\quad \text{ETHNIC COMPOSITION}) + \text{R5}\end{aligned}$$

Stable centrality (P5, B52) moderated the daily-level relation between being surrounded by same-ethnic others and ethnic private regard (Table 3). As hypothesized, simple slopes tests found that for Asian adolescents who reported that their ethnic identity was important to them (top 15%), being surrounded by more Asians was associated with feeling more positive about being a member of their ethnic group ($b = 0.73$, $SE = .09$, $p < .05$; Figure 2). In contrast, Asian adolescents who reported lower levels of ethnic identity centrality (lowest 15%) showed a negative daily-level association between being surrounded by same-ethnic others and private regard ($b = -1.73$, $SE = .06$, $p < .01$). For Asian adolescents who reported that their ethnic identity was not important, being surrounded by more Asians was associated with feeling worse about being a member of their ethnic group.

In contrast to being surrounded by other Asians, there was no moderating effect of centrality on interacting with same-ethnic others. Consistent with results found for Level 1, interacting with same-ethnic

Table 3

Hierarchical Linear Models Estimates of Daily Intragroup Contact Predicting Daily Private Regard Moderated by Stable Ethnic Identity Centrality and School Ethnic Composition

		B	*SE*
Daily private regard	B00, G000	1.31	0.62
Stable private regard	B01, G010	0.80	0.23
Stable centrality	B02, G020	−0.43	0.18
Self-reported school comp.	B03, G030	0.36	0.95
Self-reported neigh. comp.	B04, G040	0.35	0.80
Day of study	B10, G100	−0.03	0.02
Day type	B20, G200	0.04	0.12
Daily centrality	B30, G300	0.62***	0.08
Contact—interacted	B40, G400	0.26	0.33
Stable private regard	B41, G410	−0.47	0.75
DOE-reported school comp.	B41, G411	0.70	1.08
Stable centrality	B42, G420	0.61	0.62
DOE-reported school comp.	B42, G421	−1.03	0.90
Self-reported school comp.	B43, B430	−0.30	0.52
Self-reported neigh. comp.	B44, B440	−0.04	0.28
Contact—surrounded	B50, G500	−0.38	0.31
Stable private regard	B51, G510	1.40	0.79
DOE-reported school comp.	B51, G511	−2.17	1.15
Stable centrality	B52, G520	−1.39*	0.66
DOE-reported school comp.	B52, G521	2.38*	0.97
Self-reported school comp.	B53, G530	0.45	0.40
Self-reported neigh. comp.	B54, G540	0.04	0.31

Note. Using methods described in Xu (2003), the R^2 for the above model is estimated to be 0.37, such that 37% of the observed variability in daily private regard is explained by the variables included in the model. School comp. = school ethnic composition; Neighborhood comp. = neighborhood ethnic composition; DOE = Department of Education.
*$p < .05$. ***$p < .001$.

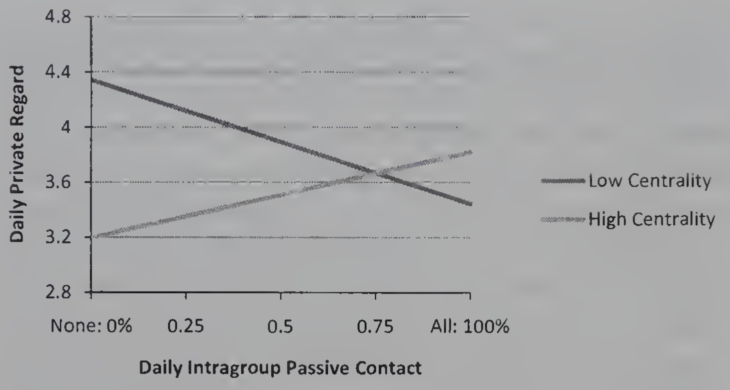


Figure 2. Interaction between stable ethnic identity centrality and being surrounded by same-ethnic others predicting daily private regard.

others was not related to daily private regard (Table 3).

School Ethnic Composition as a Moderator

Building upon the previous model, we included objective school ethnic composition at Level 3 to examine if it moderated the relation between stable centrality and interacting with same-ethnic others on private regard (B42, G421; Figure 1, Path 5) as well as moderated the relation between stable centrality and being surrounded by same-ethnic others on private regard (B52, G521; Figure 1, Path 6). Results revealed a significant three-way interaction between structural (i.e., school ethnic composition), personal (i.e., stable ethnic centrality), and daily (i.e., surrounded by same-ethnic others) factors for ethnic private regard (Table 3). For clarity, we present the simple slopes for this interaction by school context (i.e., predominantly Asian, predominantly White, and heterogeneous).

Predominantly Asian school. For Asian adolescents attending a predominantly Asian school, there was no daily-level relation between being surrounded by same-ethnic others and private regard regardless of stable centrality (Figure 3a); simple slopes tests found that the slopes for high and low centrality were not significantly different from zero. Therefore, there does not seem to be an association between daily passive intragroup contact and private regard for Asian adolescents attending a predominantly Asian school. In addition, there was no difference between the intercepts for the high- and low-centrality adolescents, suggesting that there were no differences in private regard according to centrality.

Predominantly White school. For Asian adolescents attending the predominantly White school, centrality influenced the daily-level relation between being surrounded by same-ethnic others and private regard (Figure 3b). Specifically, for Asian adolescents who reported that their ethnic identity was important to them, on days when they were surrounded by more Asians, the better they felt about being a member of their ethnic group ($b = 0.04$, $SE = .02$, $p < .05$). However, for Asian adolescents who reported that their ethnic identity was not important to them, the more Asians they

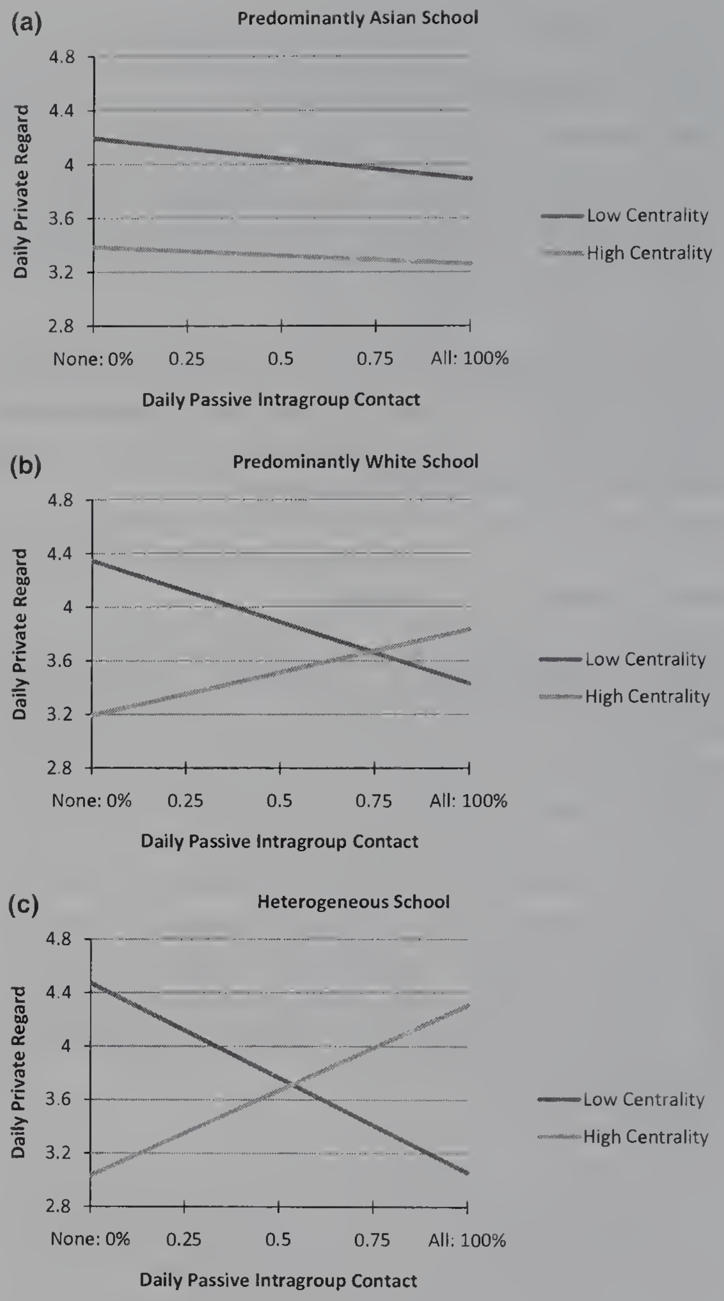


Figure 3. (a) The moderating effect of ethnic identity centrality on the daily-level relation between being surrounded by same-ethnic others and private regard in the predominantly Asian school. (b) The moderating effect of ethnic identity centrality on the daily-level relation between being surrounded by same-ethnic others and private regard in the predominantly White school. (c) The moderating effect of ethnic identity centrality on the daily-level relation between being surrounded by same-ethnic others and private regard in the heterogeneous schools.

were surrounded by, the worse they felt about being a member of their ethnic group ($b = -0.35$, $SE = .01$, $p < .01$).

Heterogeneous school. For Asian adolescents attending heterogeneous schools, centrality influenced the daily-level relation between being surrounded by same-ethnic others and private regard similar to patterns observed in predominantly White schools (Figure 3c). Specifically, for Asian adolescents who reported that their ethnic identity was important to them, the more Asians they were surrounded by, the better they felt about being a member of their ethnic group ($b = 0.17$, $SE = .02$, $p < .05$). However, for Asian adolescents who reported that their ethnic identity was not important to them, the more Asians they were surrounded by, the worse they felt about being a member of their ethnic group ($b = -0.20$, $SE = .01$, $p < .01$).

In contrast to being surrounded by same-ethnic others, there was no moderating effect of school composition on the relation between stable centrality and interacting with same-ethnic others on private regard. Consistent with all previous results, active contact did not predict daily private regard (Table 3).

Establishing Directionality

Having established that the daily-level association between being surrounded by same-ethnic others and private regard varies by adolescents' centrality and school ethnic composition, we turned to the directionality of this relation, which will address the degree to which adolescents demonstrate agency in selecting their daily transient contexts. We used HLM time lag analyses to examine if previous-day (d) experiences (d − 1) predict current-day outcomes.

Table 4

Hierarchical Linear Models Estimates of Previous-Day (d) Intragroup Contact Predicting Daily Private Regard Moderated by Stable Ethnic Identity Centrality and School Ethnic Composition

		B	*SE*
Daily private regard, d	B00, G000	−0.68	0.85
Stable private regard	B01, G010	0.42	0.40
Stable centrality	B02, G020	−0.16	0.34
Self-reported school comp.	B03, G030	1.35	1.12
Self-reported neigh. comp.	B04, G040	1.05	1.39
Day of study	B10, G100	−0.01	0.03
Day type	B20, G200	0.09	0.11
Daily centrality, d	B30, G300	0.59**	0.10
Contact—same-day interacted, d	B40, G400	0.03	0.08
Contact—same-day surrounded, d	B50, G500	−0.08	0.06
Contact—previous-day interacted, day − 1	B60, G600	0.17	0.31
Stable private regard	B61, G610	0.06	0.80
DOE-reported school comp.	B61, G611	0.05	1.17
Stable centrality	B62, G620	0.76	0.69
DOE-reported school comp.	B61, G611	−1.20	1.01
Self-reported school comp.	B63, G630	−0.24	0.44
Self-reported neigh. comp.	B64, G640	−0.08	0.29
Contact—previous-day surrounded, d −1	B70, G700	−0.01	0.32
Stable private regard	B71, G710	0.59	0.85
DOE-reported school comp.	B71, G711	−1.04	1.24
Stable centrality	B72, G720	−2.70***	0.79
DOE-reported school comp.	B72, G721	4.16***	1.18
Self-reported school comp.	B73, G730	0.26	0.55
Self-reported neigh. comp.	B74, G740	−0.10	0.34
Previous-day private regard, d − 1	B80, G800	0.23	0.14

Note. Using methods described in Xu (2003), the R^2 for the above model is estimated to be 0.49, such that 49% of the observed variability in daily private regard is explained by the variables included in the model. School comp. = school ethnic composition; Neighborhood comp. = neighborhood ethnic composition; DOE = Department of Education.
$^{**}p < .01$. $^{***}p < .001$.

Predicting Private Regard

We included the same control variables as those in the previous set of concurrent-day analyses. First, current-day private regard was modeled as a function of being surrounded by and interacting with same-ethnic others on the previous day and previous-day private regard (d − 1), controlling for same-day contact. We tested interacting with same-ethnic others in these models because although the concurrent-day models did not reveal significant effects, it is possible that the lagged analyses would. There were no significant relation between the previous-day intragroup contact variables (interacted with [B60; Figure 1, Path 1] and surrounded by [B70; Figure 1, Path 2]) and private regard (Table 4). Second, stable centrality was entered as a Level 2 moderating variable (P6, B62; P7, B72). Because previous-day (d − 1) contact was of central interest in these analyses, we entered all Level 2 variables as predictors of previous day intragroup contact, while same-day intragroup contact was controlled (interacted with, B40; surrounded by, B50). Mirroring concurrent-day results, stable centrality (P7, B72; Figure 1, Path 4) moderated the daily-level relation between being surrounded by same-ethnic others on the previous day and private regard. Specifically, simple slope analyses indicated that for Asian adolescents who reported that their ethnic identity was important to them, the more they were surrounded by same-ethnic others the previous day, the better they felt about being a member of their ethnic group the next day ($b = 2.32$, $SE = .01$, $p < .01$). However, for Asian adolescents who reported that their ethnic identity was not important to them, the more they were surrounded by same-ethnic others the previous day, the worse they felt about being a member of their ethnic group the next day ($b = -1.56$, $SE = .07$, $p < .01$). Finally, we included objective school ethnic composition as a Level 3 indicator (B42, G421; Figure 1, Path 5 and B52, G521; Figure 1, Path 6). Results revealed a significant three-way cross-level interaction that was consistent with the results from the concurrent-day models. For clarity, corresponding simple slopes results are presented by school context.

Predominantly Asian school. For Asian adolescents attending school with a predominantly Asian population, simple slopes test revealed no significant relation between being surrounded by same-ethnic others the previous day and next-day private regard, regardless of students' level of centrality. The slopes for high and low centrality were not significantly different from zero, and the intercepts were not different from each other.

Predominantly White school. For Asian adolescents attending a predominantly White school, centrality influenced the relation between being surrounded by same-ethnic others the previous day and private regard the next day. Specifically, for Asian adolescents who reported that their ethnic identity was important to them, the more they were surrounded by Asians the previous day, the better they felt about being Asian the next day ($b = 0.13$, $SE = .01$, $p < .01$). However, for Asian adolescents who reported that their ethnic identity was not important to them, the more they were surrounded by Asians the previous day, the worse they felt about being Asian the next day ($b = -0.10$, $SE = .01$, $p < .01$).

Heterogeneous school. For Asian adolescents attending an ethnically heterogeneous school, centrality influenced the relation between being surrounded by same-ethnic others the previous day and private regard the next day. Specifically, for Asian adolescents who reported that their ethnic identity was important to them, the more they were surrounded by Asians the previous day, the better they felt about being Asian the next day ($b = 0.40$, $SE = .01$, $p < .01$). However, for Asian adolescents who reported that their ethnic identity was not important to them, the more they were surrounded by Asians the previous day, the worse they felt about being Asian the next day ($b = -0.10$, $SE = .01$, $p < .01$).

Predicting Intragroup Contact

To determine whether or not ethnic private regard predicts subsequent day intragroup contact, we modeled current-day intragroup contact as a function of previous-day private regard and previous-day contact (d − 1), controlling for same-day private regard and intragroup contact (d). We conducted one model for contact measured as surrounded by same-ethnic others and one model for contact measured as interacting with same-ethnic others (Figure 4). We used the same control variables as in previous models.

Surrounded by same-ethnic others. First, there was no significant relation between previous-day private regard and being surrounded by same-ethnic others the next day (Figure 4, Path 1). Second, we entered stable centrality as a Level 2 moderating variable (Figure 5, Path 2). Because previous-day (d − 1) private regard is the predictor of interest in the current model, we entered all Level 2 variables as predictors of previous-day private regard (P4). Unlike previous models, none of the Level 2 models moderated the association between private regard the previous day and being surrounded by other Asians the next day. Finally, we included objective school ethnic composition at Level 3 (Figure 4, Path 3); however, there was no significant effect. These null findings suggest that the daily-level relation between being surrounded by other Asians and private regard is unidirectional, specifically, that being surrounded by Asians the previous-day predicts private regard the next day rather than the reverse.

Interacting with same-ethnic others. First, there was no significant relation between private regard the previous day and interacting with Asians the next day (Figure 4, Path 1). Second, we entered

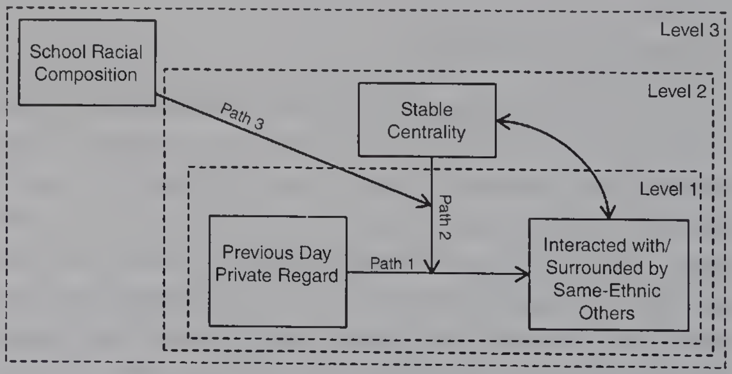

Level 3
School Racial Composition
Level 2
Path 3
Stable Centrality
Level 1
Path 2
Previous Day Private Regard
Path 1
Interacted with/ Surrounded by Same-Ethnic Others

Figure 4. Schematic for multilevel model using previous-day private regard to predict interacting with same-ethnic others or being surrounded by same-ethnic others, as moderated by stable centrality and school ethnic composition. Parallel but independent models were run for interacting with and being surrounded by same-ethnic others.

stable centrality as a Level 2 moderating variable (Figure 4, Path 2). As previous-day (d − 1) private regard is the predictor of interest in the current model, we entered all Level 2 variables as moderators of previous-day private regard (P4). None of the Level 2 variables moderated the daily-level associations. Finally, we included objective school ethnic composition at Level 3 (Figure 4, Path 3); however, there was no significant effect. Together, these null findings suggest that Asian adolescents do not demonstrate agency in interacting with same-ethnic others today based on the previous day's feelings about their ethnic group.

Discussion

We employed ecological and self-categorization theories (Turner et al., 1987) to examine how contact with same-ethnic others, ethnic centrality, and school ethnic composition shape Asian adolescents' feelings about their ethnic group. We found that Asian adolescents' daily ethnic private regard varied by personal and structural factors. In schools where the majority of peers were White or where there was no single majority ethnic population, Asian adolescents who were highly identified with their ethnic group reported feeling better about their ethnic group on days that they were surrounded by other Asians. However, in these same school settings, Asian adolescents who were less identified with their ethnic group reported feeling worse about their ethnic group on days that they were surrounded by other Asians. In addition, being surrounded by same-ethnic others had a carry-over effect on ethnic identity for Asians, who highly identified with their ethnic group. Specifically, the more Asians they were surrounded by today, the more positively they felt about their ethnic group the next day. Collectively, these findings highlight the importance of studying the confluence of adolescents' daily experiences across contexts, and they contribute to work on how distal and proximal contexts influence ethnic identity development over time (e.g., Tsai & Fuligni, 2012).

Not only did we observe that intragroup contact has implications for how Asian adolescents feel about their ethnic group but we also note that the form of intragroup contact matters. Recognizing that intragroup contact has been conceptualized and operationalized in multiple ways—simply being surrounded by same-ethnic others and actually interacting with same-ethnic others—we included both constructs. One major contribution of our study is that we found that simply having same-ethnic others around is more influential than interacting with them. This is particularly the case for Asian adolescents who reported that their ethnic identity was important to them and attended predominantly White or ethnically heterogeneous schools. This finding is consistent with existing research that shows that the presence of similar others is comforting and associated with more positive self-perceptions (Zhou, 2006), particularly when one's group is in the minority (Frable et al., 1998). When similar others are not typically available, their presence may be associated with increased self-perception because similar others generally have a more positive view of other in-group members and are able to nurture one's ethnic identity (Frable et al., 1998). Interestingly, we found consistent effects for the mere presence of same-ethnic others, but no effects for more involved interactions. It is possible that social comparison dynamics (Festinger, 1954) may explain this difference; when individuals are engaged in active intragroup contact, there is an opportunity for the interaction to be positive (i.e., identity affirming) or negative (i.e., identity rejecting). Passive contact (e.g., just seeing another Asian person in the room), however, is likely valence neutral because social comparison dynamics may be less relevant. For Asian adolescents who are in the minority in their school, the mere presence of same-ethnicity others seems to provide comfort and positivity.

Ethnic identity centrality has been found to moderate a variety of associations (Rowley, Sellers, Chavous, & Smith, 1998; Shelton & Sellers, 2000). Consistent with this previous study, we found that ethnic centrality is an important individual difference in adolescents' daily lives. Although existing research shows that contact with similar others is associated with positive self-perceptions (Frable et al., 1998), we found that this relation was only true for adolescents who were highly identified with their ethnic group. For adolescents who reported that being Asian was important to their identity, being surrounded by Asians was associated with feeling good about being Asian. In contrast, for Asian adolescents who reported that being Asian was less important, being surrounded by Asians was associated with feeling worse about being Asian. Contact with other Asians only conferred positive benefits to the ethnic feelings of adolescents who valued their ethnic identity, a finding consistent with self-categorization theory's notion that identity-relevant environmental cues (i.e., intragroup contact) have more impact on individuals who value that identity (Turner et al., 1987).

Consistent with developmental theorists who underscore the importance of youths' daily contexts (Barker, 1968; Gallimore, Goldenberg, & Weisner, 1993), we found that school ethnic context played a significant role in how Asian adolescents daily feelings about their ethnic group. As expected, schools in which Asian adolescents were in the numerical majority differed from those in which they were not. Adolescents in predominantly Asian schools reported no daily-level association between intragroup contact and private regard, irrespective of individual differences in centrality. Perhaps for these adolescents, being surrounded by other Asian adolescents is the "norm" and therefore did not have any bearing on their feelings about being Asian. Or, as self-categorization theory (Turner et al., 1987) would predict, personal attributes are more salient than group-level identities when adolescents are in predominantly in-group settings. In support of this hypothesis, although intragroup contact has typically been portrayed as positive, this is not always the case. Research on ethnic discrimination finds that Asian American adolescents are particularly likely to be targeted by in-group members (Fisher et al., 2000; Rivas-Drake et al., 2008). Asians comprise a diverse group that includes over 30 different ethnic groups (Espiritu, 1992); coupled with differences in language, generational status, and socioeconomic status, there are many opportunities to create distinctions within the group. Consistent with social comparison theory (Festinger, 1954) and as observed in an ethnographic study (Lee, 1996), interacting with same-ethnic others includes both negative and positive exchanges. We also note that although some of the Asian adolescents in our sample attended schools in which they were in the majority, in the larger U.S. context, Asians comprise only 5% of the population.

In contrast, Asian adolescents who were in predominantly White schools or schools with no clear ethnic majority were likely to find being surrounded by other Asians a relatively rare occurrence. As such, on days in which Asian adolescents at these schools were around other Asians, they reported differences in private regard based on feelings of centrality. Specifically, for adolescents who reported that being Asian was important to their identity, being surrounded by Asians was associated with feeling good about being Asian. In contrast, for Asian adolescents who reported that being Asian was less important to them, being surrounded by Asians was associated with feeling worse about being Asian. Asian adolescents whose ethnicity is not important to them are likely to have other identities (e.g., religion) that are more integral to their identity. When one makes a conscious decision to make ethnicity not important to one's overall identity, the presence of other Asians may be associated with lower private regard because the in-group serves as a reminder of that decision. Therefore, students may experience the same setting in different ways depending on the attachment to their ethnic identity.

The literature on ethnic identity development has characterized adolescents as "recipients" and "agents" of socializing forces such as parents, peers, schools, and neighborhoods (Hughes & Chen, 1997; Seaton & Yip, 2009). We conducted time lagged analyses to examine if adolescents' daily ethnic private regard fluctuated according to socializing forces in their immediate context. Specifically, we explored whether the amount and type of intragroup contact adolescents had today predicted their ethnic private regard the next day, or whether ethnic private regard today predicted the amount and type of intragroup contact the next day. Although the former suggests that adolescents seek to affiliate with in-group members as a consequence of their positive ethnic feelings, the latter suggests that adolescents are responsive to their environments. We found that the more adolescents were surrounded by same-ethnic others today, the more positively they felt about their ethnic group the next day; the reverse, however, was not true. Notably, this effect was not found for how much adolescents interacted with same-ethnic others, suggesting that simply being surrounded by other Asians was sufficient to influence private regard the next day. Daily intragroup contact has been theorized as predicting subsequent feelings of ethnic identity (Levin et al., 2003; Tatum, 1997), and we found empirical support for this. Although our data paint a portrait of adolescents who are responding to their environments rather than creating them, we propose a few caveats. First, we observed results for passive, not active, contact and it may be more difficult for adolescents to exert control over who surrounds them. Second, our data represent ninth-grade adolescents; it is possible that as adolescents develop, they exert more agency over their environments. Finally, we only examined the associations between private regard and intragroup contact; it is possible that examining other variables, such as friendship affiliation, would shed more light on how adolescents affect their ethnic identity development.

Although this study makes a significant contribution to the understanding of adolescent ethnic identity, limitations remain. For one, although we examined the quantity of contact with same-ethnic others, we did not examine the quality of this contact and we believe that this is an important area for future research. On a related note, while no participants indicated having questions or concerns about answering questions about contact with "same-ethnicity" others, as suggested by self-categorization theory's notion of *comparative context* (Turner et al., 1987), the interpretation of the question could vary across contexts. For example, in situations where Asians are in the minority, "sameness" might involve the inclusion of multiple Asian ethnicities to form a pan-ethnic category (Espiritu, 1992). However, in settings that are predominantly Asian and include a diversity of Asian ethnicities, definitions of "sameness" may only include individuals who are the same ethnicity (Lee, 1996). Because the question did not ask participants to differentiate between race and ethnicity, we are unable to disentangle such differences. Moreover, while our data represent a dynamic snapshot of ethnic identity development, to truly understand the development of stable ethnic identity it would be important to explore these associations longitudinally. We may find, for example, that the daily pairing of intragroup contact and private regard has implications for how adolescents develop a sense of stable centrality or private regard 1 year later, thereby directly linking daily experiences to longer term developmental process. It seems plausible that adolescents who experience daily increases in private regard may, over time, develop a stronger attachment to their ethnic identity. In addition, future research would also benefit from exploring other contexts that influence adolescent development (i.e., study, church) that may vary in diversity. As this article includes only an Asian sample, we do not know if the same patterns would be observed in other groups or between Asian subgroups. Also, our predominantly female sample makes it difficult to observe possible gender effects. Finally, we call attention to possible measurement issues. We adapted the Centrality and Private Regard subscales from the MIBI (Sellers et al., 1997), which was developed for and normed on African American samples. "Black" from the original scale was replaced with "your ethnic or racial group"; in doing so, lower reliability was observed for the centrality subscale than has typically been found among African American samples.

Concluding Thoughts

Considering daily intragroup contexts, school ethnic composition, and individual differences in ethnic identity, this study highlights the importance of examining the confluence of these factors in understanding adolescents' everyday experiences of ethnicity and race. Coupled with existing cross-sectional and longitudinal data, we can begin to develop a richer and more nuanced understanding of the role of ethnicity and race in adolescents' daily lives. In doing so, we move away from conceptualizing and operationalizing ethnicity and race as static characteristics of adolescents, but rather as a lens through which everyday experiences are lived.

References

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin*, *130*, 80–114.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, *1*, 164–180.

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.

Branscombe, N. R., Deaux, K., & Lerner, M. S. (1985). Individual differences and the influence of context on categorization and prejudice. *Representative Research in Social Psychology*, *15*, 25–35.

Branscombe, N. R., Spears, R., Ellemers, N., & Doosje, B. (2002). Intragroup and intergroup evaluation effects on group behavior. *Personality and Social Psychology Bulletin*, *28*, 744–753.

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist*, *62*, 728–738.

Bronfenbrenner, U. (Ed.) (1999). *Environments in developmental perspective: Theoretical and operational models*. Washington, DC: American Psychological Association.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Byrd, C. M., & Chavous, T. M. (2009). Racial identity and academic achievement in the neighborhood context: A multilevel analysis. *Journal of Youth and Adolescence*, *38*, 544–559.

Chavous, T. M., Harris, A., Rivas, D., Helaire, L., & Green, L. (2004). Racial stereotypes and gender in context: African Americans at predominantly Black and predominantly White colleges. *Sex Roles*, *51*, 1–16.

Erikson, E. (1968). *Identity: Youth and crisis*. New York: Norton.

Espiritu, Y. L. (1992). *Asian American panethnicity: Bridging institutions and identities*. Philadelphia, PA: Temple University Press.

Ethier, K. A., & Deaux, K. (1994). Negotiating social identity when contexts change: Maintaining identification and responding to threat. *Journal of Personality and Social Psychology, 67*, 243–251.

Ethier, K. A., & Deaux, K. (2001). Negotiating social identity when contexts change: Maintaining identification and responding to threat. In M. A. Hogg (Ed.), *Intergroup relations: Essential readings. Key readings in social psychology* (pp. 254–265). Philadelphia, PA: Psychology Press.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.

Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence, 29*, 679–695.

Forehand, M. R., Deshpande, R., & Reed, A., II (2002). Identity salience and the influence of differential activation of the social-schema on advertising response. *Journal of Applied Psychology, 87*, 1086–1099.

Frable, D. E., Platt, L., & Hoey, S. (1998). Concealable stigmas and positive self-perceptions: Feeling better around similar others. *Journal of Personality and Social Psychology, 74*, 909–922.

Gallimore, R., Goldenberg, C. (1993). Activity settings of early literacy: Home and school factors in children's emergent literacy. In E. A. Forman, N. Minick & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 315–335). Oxford, England: Oxford University Press.

Gallimore, R., Goldenberg, C. N., & Weisner, T. S. (1993). The social construction and subjective reality of activity settings: Implications for community psychology. *American Journal of Community Psychology, 21*, 537–559.

Gurin, P., & Townsend, A. (1986). Properties of gender identity and their implications for gender consciousness. *British Journal of Social Psychology, 25*, 139–148.

Hamm, J. V. (2000). Do birds of a feather flock together? The variable bases for African American, Asian American, and European American adolescents' selection of similar friends. *Developmental Psychology, 36*, 209–219. doi:10.1037/0012-1649.36.2.209

Hartup, W. W. (2005). Peer Interaction: What causes what? *Journal of Abnormal Child Psychology, 33*, 387–394.

Hughes, D., & Chen, L. (1997). When and what parents tell children about race: An examination of race-related socialization among African American families. *Applied Developmental Science, 1*, 200–214.

Hutnik, N., & Sapru, S. (1996). The salience of ethnicity. *Journal of Social Psychology, 136*, 661–662.

Juvonen, J., Nishina, A., & Graham, S. (2006). Ethnic diversity and perceptions of safety in urban middle schools. *Psychological Science, 17*, 393–400.

Lee, R. M., Noh, C.-Y., Yoo, H. C., & Doh, H.-S. (2007). The psychology of diaspora experiences: Intergroup contact, perceived discrimination, and the ethnic identity of Koreans in China. *Cultural Diversity and Ethnic Minority Psychology, 13*, 115–124. doi:10.1037/1099-9809.13.2.115

Lee, S. J. (1996). *Unraveling the "model minority" stereotype: Listening to Asian American youth*. New York: Teachers College Press.

Levin, S., van Laar, C., & Sidanius, J. (2003). The effects of ingroup and outgroup friendship on ethnic attitudes in college: A longitudinal study. *Group Processes & Intergroup Relations, 6*, 76–92.

Martin, A., & Hetrick, E. (1988). The stigmatization of the gay and lesbian adolescent. *Journal of Homosexuality, 15*, 163–183.

Martin, P. P., Wout, D., Nguyen, H., Sellers, R. M., & Gonzalez, R. (2005). *Investigating the psychometric properties of the Multidimensional Inventory of Black Identity in two samples: The development of the MIBI-S*. Unpublished manuscript.

Matschke, C., & Sassenberg, K. (2010). Does rejection lead to disidentification? The role of internal motivation and avoidance strategies. *European Journal of Social Psychology, 40*, 891–900.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

Neal-Barnett, A., Stadulis, R., Singer, N., Murray, M., & Demmings, J. (2010). Assessing the effects of experiencing the acting White accusation. *The Urban Review, 42*, 102–122.

Niemann, Y. F., & Dovidio, J. F. (1998). Relationship of solo status, academic rank, and perceived distinctiveness to job satisfaction of racial/ethnic minorities. *Journal of Applied Psychology, 83*(1), 55–71.

Niemann, Y. F., Pollak, K. I., Rogers, S., & O'Connor, E. (1998). Effects of physical context on stereotyping of Mexican American males. *Hispanic Journal of Behavioral Sciences, 20*, 349–362.

O'Donnell, C. R., Tharp, R. G., & Wilson, K. (1993). Activity settings as the unit of analysis: A theoretical basis for community intervention and development. *American Journal of Community Psychology, 21*, 501–520.

Phinney, J. S. (1993). A three-stage model of ethnic identity development in adolescence. In M. E. Bernal (Ed.), *Ethnic identity: Formation and transmission among Hispanics and other minorities* (pp. 61–79). Albany: State University of New York Press.

Phinney, J. S., Romero, I., Nava, M., & Huang, D. (2001). The role of language, parents, and peers in ethnic identity among adolescents in immigrant families. *Journal of Youth and Adolescence, 30*, 135–153.

Rivas-Drake, D., Hughes, D., & Way, N. (2008). A closer look at peer discrimination, ethnic identity, and psychological well-being among urban Chinese American sixth graders. *Journal of Youth and Adolescence, 37*, 12–21.

Rowley, S. J., Sellers, R. M., Chavous, T. M., & Smith, M. A. (1998). The relationship between racial identity and self-esteem in African American college and high school students. *Journal of Personality and Social Psychology, 74*, 715–724.

Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition, 12*, 61–74.

Sanchez, D. T., & Garcia, J. A. (2009). When race matters: Racially stigmatized others and perceiving race as a biological construction affect biracial people's daily well-being. *Personality and Social Psychology Bulletin, 35*, 1154–1164. doi:10.1177/0146167209337628

Sanchez, D. T., Shih, M., & Garcia, J. A. (2009). Juggling multiple racial identities: Malleable racial identification and psychological well-being. *Cultural Diversity and Ethnic Minority Psychology, 15*, 243–254.

Seaton, E. K., & Yip, T. (2009). School and neighborhood contexts, perceptions of racial discrimination, and psychological well-being among African American adolescents. *Journal of Youth and Adolescence, 38*, 153–163.

Seder, J. P., & Oishi, S. (2009). Ethnic/racial homogeneity in college students' facebook friendship networks and subjective well-being. *Journal of Research in Personality, 43*, 438–443.

Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997). Multidimensional Inventory of Black Identity: A preliminary investigation of reliability and construct validity. *Journal of Personality and Social Psychology, 73*, 805–815.

Sellers, R. M., Shelton, J. N., Cooke, D. Y., Chavous, T. M., Rowley, S. A. J., & Smith, M. A. (Eds.) (1998). *A multidimentional model of racial identity: Assumptions, findings, and future directions*. Hampton, VA: Cobb & Henry.

Sellers, R. M., Smith, M. A., Shelton, J. N., Rowley, S. A. J., & Chavous, T. M. (1998). Multidimensional model of racial identity: A reconceptualization of African American racial identity. *Personality and Social Psychology Review*, 2, 18–39.

Shelton, J. N., & Sellers, R. M. (2000). Situational stability and variability in African American racial identity. *Journal of Black Psychology, 26*, 27–50.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development, 9*, 545–569.

Tatum, B. D. (1997). *"Why are all the Black kids sitting together in the cafeteria?" and other conversations about race*. New York: Basic Books.

Tatum, B. D. (2004). Family life and school experience: Factors in the racial identity development of Black youth in White communities. *Journal of Social Issues, 60*, 117–135.

Tsai, K. M., & Fuligni, A. J. (2012). Change in ethnic identity across the college transition. *Developmental Psychology, 48*, 56–64. doi:10.1037/a0025678

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research*. Greenwich, CT: JAI Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Uba, L. (1994). *Asian Americans: Personality patterns, identity, and mental health*. New York: Guilford.

Umaña-Taylor, A. J. (2004). Ethnic identity and self-esteem: Examining the role of social context. *Journal of Adolescence, 27*, 139–146. doi:10.1016/j.adolescence.2003.11.006

U.S. Census Bureau. (2010). 2010 census shows Asians are fastest-growing race group. Retrieved October 30, 2011, from http://2010.census.gov/news/releases/operations/cb12-cn22.html

Whiting, J., & Whiting, B. (1975). *Children of six cultures: A psychocultural analysis*. Cambridge, MA: Harvard University Press.

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22*, 3527–3541.

Yip, T. (2005). Sources of situational variation in ethnic identity and psychological well-being: A Palm Pilot study of Chinese American students. *Personality and Social Psychology Bulletin, 31*, 1603–1616.

Yip, T., & Fulgni, A. J. (2002). Daily variation in ethnic identity, ethnic behaviors, and psychological well-being among American adolescents of Chinese descent. *Child Development, 73*, 1557–1572. doi:10.1111/1467-8624.00490

Yip, T., Seaton, E. K., & Sellers, R. M. (2010). Interracial and intraracial contact, school-level diversity, and change in racial identity status among African American adolescents. *Child Development, 81*, 1431–1444. doi:10.1111/j.1467-8624.2010.01483.x

Zhou, M. (2006). Negotiating culture and ethnicity: Intergenerational relations in Chinese immigrant families in the United States. In R. Mahalingam (Ed.), *Cultural psychology of immigrants* (pp. 315–336). Mahwah, NJ: Erlbaum.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1442–1458

Siblings, Theory of Mind, and Executive Functioning in Children Aged 3–6 Years: New Longitudinal Evidence

Anna R. McAlister
Michigan State University

Candida C. Peterson
University of Queensland

Longitudinal data were obtained from 157 children aged 3 years 3 months to 5 years 6 months at Time 1. At Time 2 these children had aged an average of 12 months. Theory of mind (ToM) and executive functioning (EF) were measured at both time points. Results suggest that Time 1 ToM scores predict Time 2 EF scores. Detailed examination of sibling influences suggests that benefits—in terms of advanced ToM development—accrue to children with siblings versus without, and to those with a larger number of child-aged siblings. Any advance in either area (ToM or EF) is likely to benefit the other, and early sibling interaction appears to act as a catalyst.

Cognitive developments of the preschool period are impressive (Flavell, 2004) and two of the most notable of these are theory of mind (ToM) and executive functioning (EF). ToM involves a representational understanding of others' minds including abilities to comprehend and predict others' mental states of true and false belief, memory, imagination, and the like, even in situations where these mental states are at odds with observable reality. Developing dramatically over the period from ages 2 to 6, the prototypical tests for ToM (i.e., inferential false belief tasks) are failed by most preschoolers at age 3 but are mastered so universally by age 5 as to suggest that understanding of belief and of mind exhibit genuine conceptual change in the preschool years (Wellman, Cross, & Watson, 2001). Whereas some theoretical accounts (e.g., Leslie & Thaiss, 1992) ascribe ToM development to the maturational unfolding of a preprogrammed, neurocognitive module largely independent of variations in familial, social, or educational experiences, many others highlight individual differences among preschoolers in the pace of ToM growth (e.g., Dunn & Brophy, 2005). Evidence also suggests that faster rates of ToM progress among typically developing children are linked with variations in social circumstances (e.g., number of siblings; Perner, Ruffman, & Leekam, 1994) and life experience (e.g., richness of maternal mental state conversation; Ruffman, Slade, & Crowe, 2002), suggesting possible causal connections (Dunn & Brophy, 2005).

Another key development of the preschool years involves increases in the EF skills that control, direct, and monitor cognitive function and its behavioral outcomes (Zelazo, Carlson, & Kesek, 2008). These include inhibitory control, working memory, resistance to interference, and set shifting (Garon, Bryson, & Smith, 2008). EF skills are believed to be responsible for conscious, effortful control of thoughts and actions (Oh & Lewis, 2008), and are distinct from intelligence, including both verbal and nonverbal IQ. Cross-sectional links between preschoolers' ToM and EF skills are well established (Moses, Carlson, & Sabbagh, 2006).

Several distinct theoretical accounts have been put forward to explain why the EF–ToM association arises (Carlson, Moses, & Breton, 2002). The expression account argues that the link is relatively superficial. Children simply need EF to display their already-developed ToM understanding in testing situations, especially those relying on inferential false belief tests. Alternatively, emergence accounts view the association between the two capacities as more fundamental than a simple test-performance limitation. Although varying in detail, these argue that much of the influence of EF is relevant in terms of the level of ToM competence displayed (Carlson & Moses, 2001). Without EF capability, ToM might be impossible to acquire. Conversely, EF emergence may equally be contingent upon the prior development of ToM (e.g., Perner & Lang, 1999). Without

---

Correspondence concerning this article should be addressed to Anna R. McAlister, Department of Advertising & Public Relations, 404 Wilson Road, Room 309, East Lansing MI 48824. Electronic mail may be sent to annamc@msu.edu.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0023
DOI: 10.1111/cdev.12043

ToM-based awareness of the mind's role and its limitations, children might not master (or apply) EF skills to the strategic control of their overt behavior. Still a third possibility is that a common mediating factor (e.g., language development) might underpin the growth of both ToM and EF during the preschool period, thus explaining their observed correlations with one another. Others have suggested that a child gains better self-control once he or she has a better understanding of his or her own mind (Perner, 1991; Perner, Lang, & Kloo, 2002). Moreover, some researchers entertain the notion that EF and ToM share a common set of representational abilities (Perner, 1998) or common brain regions (Ozonoff, Pennington, & Rogers, 1991). The cognitive complexity and control theory (Frye, Zelazo, & Burack, 1998) posit that success on false belief tests requires a type of reasoning not unique to the understanding of mental states. Instead, the general intellectual ability to deal with abstract rules and relations is thought to underpin both ToM and EF.

Regardless of the particular mechanism at play, cross-sectional research strongly supports an overall connection between ToM and EF in young preschoolers, while highlighting a special link between ToM and inhibitory control (Carlson & Moses, 2001). However, the cross-sectional methodology has limitations when it comes to elucidating developmental pathways and distinguishing precursor, consequent, or third-variable-mediated connections. Longitudinal studies are thus desirable, particularly via the cross-lagged approach in which early and later abilities in both domains are assessed in the same children as they grow in age and experience. This enables examination of how gains made in one domain (e.g., ToM) may arise out of earlier advanced mastery of the other (e.g., EF) after controlling individual variability in initial mastery of the outcome domain itself. Accompanied by appropriate controls for other known EF and ToM correlates (like language ability), this design allows for a relatively pure assessment of pathways of causal precedence and consequence.

Cross-lagged longitudinal studies of the ToM–EF association remain scarce (Hughes & Ensor, 2007). Nonetheless, some have been conducted (e.g., Carlson, Mandell, & Williams, 2004; Farrant, Maybery, & Fletcher, 2012; Hughes, 1998; Hughes & Ensor, 2007; Müller, Liebermann-Finestone, Carpendale, Hammond, & Bibok, 2012; Schneider, Lockl, & Fernandez, 2005). Although these pioneering investigations have supplied very valuable information (see below) that has served to illuminate several key issues for the early development of EF–ToM connections, a limitation is that they have mostly concentrated on the earliest beginnings of ToM from ages 2 to 3 or early 4. There is clearly a need for additional cross-lagged longitudinal research on EF and ToM especially with somewhat older ages, as we will argue in detail later.

One recent study (Hughes & Ensor, 2007) of 2- to 4-year-olds extended earlier pioneering efforts (e.g., Hughes, 1998) to examine 122 socially diverse toddlers and found "only partial support for the view that ToM is a prerequisite for EF" (p. 1447). However, EF at age 2 significantly predicted ToM performance at ages 3 and 4. Similarly, using a microgenetic design with a small sample ($N = 21$) of 3-year-olds, Flynn, O'Malley, and Wood (2004) found that earlier EF gains were linked with later false belief mastery. However, the design entailed monthly testing over 6 months, and comparison of false belief understanding at adjacent testing times showed numerous developmental inconsistencies with children sometimes later failing tasks they previously passed. The authors therefore advocated longer term longitudinal research to identify possible broader patterns amid short-term inconsistency.

Carlson et al. (2004) sampled a group of 81 toddlers at 24 months of age and retested them at 39 months. The EF tasks these authors used showed moderate stability across this age range, but several of their lab measures of ToM at the initial test did not. In light of this, they substituted an indirect measure of ToM at 24 months, consisting of a listing by parents of all the words for internal states (perception, affect, cognition, physiology) that they believed their children knew. This language measure correlated with later scores on laboratory ToM tests, as well as with contemporaneous EF measures (at 24 months). Carlson et al. were therefore able to use the cross-lagged method to examine the predictive value of toddlers' early ToM (internal state language) on later EF performance at age 3, as well as the cross-lagged connection from early EF to later ToM. They found no clear support for the former pathway, but convincing support for the latter. Indeed, even after controlling for early ToM and language ability, children's ToM scores at 39 months were predicted by their EF skills at age 2. Similarly, Müller et al. (2012) found that EF at ages 2 and 3 predicted ToM at ages 3 and 4 even after controlling verbal ability and prior ToM. But early ToM did not influence later EF independently of language ability.

This pattern has not always emerged consistently. For example, Schneider et al. (2005) tested EF and ToM in 176 children longitudinally on three occasions, beginning at a mean age of 3.33 years and ending at 4.33 years. At each testing point, the strongest associations were between language ability and each of the other variables, with little evidence that EF and ToM were contemporaneously associated, after controlling for age and language. These findings are not fully consistent with those of Hughes and Ensor (2007) or those of Carlson et al. (2004). In fact, Schneider et al. reported no significant predictive relation between early EF performance and later ToM scores. Nor did age or working memory contribute to the prediction of ToM over and above language ability. In other words, despite their careful controls, large sample, and comprehensive batteries of ToM and EF tasks, results of this three-wave cross-lagged study were null: ToM made no unique contribution to EF, and EF made none to ToM, at any time point.

How are the inconsistencies among results of these recent, longitudinal studies of early ToM and EF connections to be reconciled? Schneider et al. (2005) suggested that one problem might have been their sample's generally poor grasp of ToM at the start of their study. In fact, throughout the longitudinal period, the standard ToM tasks they used remained difficult for their sample. Percentages passing each task did not exceed chance even at the final testing.

The rationale for studying 2- and 3-year-olds (i.e., children younger than the watershed age for false belief mastery; Wellman et al., 2001) is to capture precursors to both EF and ToM and trace their developmental relations over time. The important findings from such studies (e.g., Carlson et al., 2004; Hughes & Ensor, 2007) highlight how much can be gained from examining these constructs as early as possible in development. Nevertheless, additional research is now in order with older children (as in the present research) to discover whether early patterns are maintained as development continues.

Theoretically, there are grounds for anticipating that longitudinal connections between EF and ToM in older pre- and primary schoolers might differ from those observed in very early childhood. Peer group dynamics become more salient as children progress through pre- and primary school, posing new ToM challenges (Dunn, 1995). New EF challenges likewise arise in a more academically oriented classroom and via competitive games and sports. Indeed the known EF gains at this age imply that "by excluding school-age children we miss much of the developmental picture of EF" (Best, Miller, & Jones, 2009, p. 181), including possible changes in its patterns of interconnection with ToM. Empirically, further study of older children is desirable to balance the preponderant past focus on those younger than 4.5 years. Indeed only two known studies of older, typically developing children have been published (Farrant et al., 2012; Razza & Blair, 2009) and each had limitations. Razza and Blair (2009) failed to measure EF at Time 2, but found no predictive relation between early EF (at age 5) and ToM 1 year later. Farrant et al. (2012) failed to measure general language ability and had only one EF task. Also, paradoxically, their Time 2 ToM measure was unrelated to chronological age. Nevertheless, their findings supported several with younger preschoolers in showing early EF predicted later ToM but not vice versa. Thus, although these findings are suggestive, there is clearly a need for further study of older preschoolers and young primary schoolers, and this is one aim of the present research.

The other major aim of our study was to examine whether having siblings at home, versus being an only-child, might influence the development of ToM and EF skills and their connections to one another. Past research has suggested that preschoolers often develop ToM understanding rapidly when they have one or more siblings to interact with at home (e.g., Jenkins & Astington, 1996; Perner et al., 1994; Peterson, 2000; Ruffman, Perner, Naito, Parkin, & Clements, 1998). Children with siblings have daily access to other children's minds via mutual engagement in pretend play (Perner et al., 1994), sibling disputes (Randell & Peterson, 2009), and child-oriented conversation (Dunn & Brophy, 2005). They can also eavesdrop on parents' disciplining or teaching of siblings. Sibling interaction may also trigger strategic EF skills for winning games (Cole & Mitchell, 2000) or alert preschoolers to ToM-based consideration of others' outcomes or feelings (Ruffman, Perner, & Parkin, 1999). Thus, sibling experience might conceivably favor the rapid development of both ToM and EF. Indeed, Cole and Mitchell (2000) postulated that the development of effortful impulse control might be facilitated in households with more siblings owing to demands to delay gratification during conflict and play. Some cross-sectional evidence supports these predicted associations between the household sibling composition, EF, and ToM, independently of age and language (e.g., McAlister & Peterson, 2006). However, such associations are not always found (e.g., Cole & Mitchell, 2000).

One issue to consider in exploring links between presence of siblings and the preschooler's ToM and

EF development is the sibling's age. As Peterson (2000) noted, previous studies have not always reported information on siblings' ages and there may be upper and lower limits on the ages at which siblings may be of any developmental benefit. For example, a young infant sibling is probably too immature to interact independently with the preschooler in ways that could benefit ToM or EF development. Adolescent and adult siblings, conversely, might so resemble parents as to create a social environment similar to the only-child's. We therefore defined participants as having child-aged siblings only if the sibling was over 12 months and under 13 years in age. We then performed the following comparisons of ToM and EF performance: (a) only-child versus has at least one sibling, (b) participant's total number of child-aged siblings, and (c) birth order in the sibling constellation (eldest, etc.).

Method

Participants

The participants were 157 children (75 boys, 82 girls) aged 3 years 3 months to 5 years 6 months (M = 4 years 2 months, SD = .57 years) at Time 1. All children were recruited from Australian kindergartens and preschools that drew families from a variety of working-, middle-, and upper-middle-class neighborhoods. We had no information on the specific incomes, education, or occupations of our participants' parents. However, population census figures for the neighborhoods from which we recruited indirectly suggest that maternal education levels are predominantly high school completion or above (e.g., across the neighborhoods from which we sampled, 91% of female residents aged 20 to 34 years have completed 12 years of secondary schooling; Australian Bureau of Statistics, 2008). Most children in the sample (94.9%) came from Anglo-Australian or other English-language families (e.g., New Zealand, Ireland, Canada), whereas the remainder had at least one parent of East Asian (3.2%), Indian or Pakistani (1.3%), or German (0.6%) heritage. All children spoke English as their first language and none had any suspected or diagnosed developmental disorder. We tested all of these children again at Time 2 when they had aged an average of 12 months (SD = 1.57 months). Actual delay periods ranged from 9 months to 15 months, so that chronological ages ranged from 4 years 3 months to 6 years 8 months at Time 2 with a mean age of 5.17 years.

Sibling background. Eighty-five percent of the sample (n = 133) lived with at least one child-aged (i.e., ages 1–12 years) sibling at Time 1, whereas 24 children (15%) were only-children without child sibling access. Of these 24 singletons, 21 (or 13% of the entire sample) had no siblings at all, whereas 3 children had a preverbal infant sibling only. At Time 2, respective percentages with versus without child siblings were 87% and 13%.

Longitudinal Design

Children's ToM, EF, and language ability were individually measured at each of the longitudinal time points. At Time 1 this was at preschool. At Time 2 it also was if still enrolled, or otherwise at home (for those who had moved on to primary school). Because testing was comprehensive and lengthy, it occurred over two sessions (with a time lag of 1 week) both at Time 1 and at Time 2. The order of tasks and trials within tasks was varied systematically.

Task Order, Materials, Procedures, and Scoring

At Time 1, children completed a battery of four ToM tasks, two EF tasks, and a language measure (Peabody Picture Vocabulary Test [PPVT]). At Time 2, we selected new and more challenging test batteries for ToM (four tasks) and EF (four tasks) while repeating the PPVT. Previous empirical evidence, as detailed below, informed task selection. At Time 1, Session A had (in this order) the following: Route Navigation (two trials), Appearance–Reality Trial A, Unseen Displacement False Belief (two trials), and PPVT; whereas Session B had (in order) the following: Misleading Container False Belief, Pretend Play Trial A, Route Navigation EF (two trials), Appearance–Reality Trial B, Pretend Play Trial B, and Resisting Instructions EF. The Time 2 tasks were (Session A) as follows: Hand Game, Card Sort A, Appearance–Reality A, Card Sort B, Appearance –Reality B, Unseen Displacement False Belief (two trials), Emotion False Belief, and (Session B) Resisting Instructions EF, Real–Apparent Emotion (ToM), Tower of London EF, and PPVT. Table 1 shows all descriptive statistics.

Details and rationale for Time 1 ToM tasks. The tasks used to measure ToM at Time 1 were selected because they are widely used, reliable measures of a broad range of core ToM skills that would be likely to provide adequate variance in the level of difficulty between tasks. This was essential to detect early signs of ToM awareness among our youngest

Table 1
Descriptive Statistics for Measures at Times 1 and 2

	M	*SD*	% perfect	Minimum	Maximum
Time 1 age	4 years 2 months	0.57	—	3 years 3 months	5 years 6 months
Time 1 number of child siblings	1.34	0.87	—	0	3
Time 1 Peabody Picture Vocabulary Test (PPVT) score	47.54	16.06	—	12	82
Time 1 ToM tasks					
Unseen displacement	0.90	0.88	31.7	0	2
Misleading container	0.45	0.50	44.6	0	1
Appearance–reality	0.96	0.96	38.8	0	2
Pretend play	0.53	0.49	31.2	0	1
Time 1 ToM scale score	0.58	0.26	11.5	0	1
Time 1 EF Tasks					
Route navigation	3.51	1.36	49.8	1	4
Resisting instructions	3.24	2.64	21.7	0	6
Time 1 EF scale score	0.71	0.26	25.5	.20	1
Time 2 age	5 years 2 months	0.61	—	4 years 3 months	6 years 8 months
Time 2 number of child siblings	1.48	0.91	—	0	3
Time 2 PPVT score	67.44	18.70	—	30	113
Time 2 ToM Tasks					
Real–apparent emotion	0.18	0.39	18.5	0	1
Emotion-based false belief	1.01	0.86	36.3	0	2
Unseen displacement	0.80	0.98	29.3	0	2
Appearance–reality	1.14	1.00	40.8	0	2
Time 2 ToM scale score	0.58	0.28	12.7	0	1
Time 2 EF Tasks					
Resisting instructions	4.44	0.78	62.4	0	6
Card sort	8.26	4.06	38.9	0	14
Hand game	7.80	1.90	67.7	0	10
Tower of London	4.08	12.01	0.01	0	24
Time 2 EF scale score	0.73	0.17	1.3	0	1

participants. Because most of the children tested at Time 1 were below the age when false belief tasks are consistently passed (Wellman et al., 2001), we ensured against floor effects by following Hughes and Ensor's (2005) strategy of including tests of the understanding of the representational nature of pretend play, in addition to standard false belief and appearance–reality tasks. The two Time 1 false belief tests, unseen displacement and misleading container, had two test trials each, plus control questions. The former was a replication of Baron-Cohen, Leslie, and Frith's (1985) "Sally-Ann" task with two trials. Each included one test question and two control questions. Both were enacted and scored exactly as described by the original authors except a boy doll replaced "Ann" (for a rationale, see Peterson & Siegal, 1999). On each trial, a child could earn 1 point for correctly answering the test and all control questions. If the control questions were passed, but the test question failed, the child scored 0. Data were excluded for any child who did not pass both control questions. Taking into account performance on both trials, children could earn a maximum of 2 points for the task as a whole. The misleading container test (Gopnik & Astington, 1988) involved a Band-Aid box containing pencils. There were two test questions (one about own prior belief and one about the belief of a naïve classmate) plus one control question. To earn 1 point, children needed to pass the control question and answer both test questions correctly. Scores on unseen displacement and unexpected contents false belief tasks were significantly correlated, $r(156) = .23$, $p < .01$. Thus (so that false belief would be given equal weighting in the ToM composite to appearance–reality and pretend representation), a total false belief proportion score was created for each child by first dividing scores on the unseen displacement task by 2, then averaging across this and the misleading container task. These decimal false belief proportion scores could range from 0 to 1.

The appearance–reality test had two trials. The stimulus for the first was an apple-shaped candle. The second used a realistic flower that was actually a ballpoint pen. Each had an initial control question ("What is this?") and two test questions ("What does it look like to your eyes? and "What is it really and truly?"). Administration and scoring of each trial followed Flavell's (1986) procedure. Children passed a trial (to earn 1 point) only if all three questions were answered correctly. Trial scores were significantly correlated, $r(156) = .53$ $p < .001$, so an appearance–reality total (0–2; Table 1) was subsequently converted to a proportion score for the ToM composite.

Finally, at Time 1, awareness of the representational nature of pretending and was tested exactly as in McAlister and Peterson (2007). For each of two trials, a familiar, real object (banana or potato) was shown and participants had to pretend it was something else (i.e., the banana was a phone, the potato was soap). Once pretense was effectively enacted (which all children did), there were four test questions: (a) "What is this really?" (b) "What did we pretend this was?" Then additional objects were placed beside the stimulus, a real version of the pretend item (e.g., a real phone) and an irrelevant decoy (e.g., a tennis ball) and questions (c) "Which one did we pretend was a [phone or soap]?" and (d) "Which one is really a [phone or soap]?" were asked. Although each trial had four test questions, the maximum score per trial was 1. This was achieved only when all four questions were answered correctly. Any other response pattern earned 0. Scores on the two trials were significantly correlated, $r(156) = .24$, $p < .01$, so a proportional score was created with a range from 0 to 1.

Time 1 ToM total score. False belief proportion scores correlated with proportion scores on appearance–reality, $r(156) = .29$, $p < .001$, and pretend representation, $r(156) = .50$, $p < .001$, and the latter two were also significantly correlated, $r(156) = .29$, $p < .001$. Furthermore, each of these correlations remained statistically significant in partial correlations controlling for age and language ability, $rs(154) = .22$, .44 and .23, all $ps < .01$. Given this, a combined proportional Time 1 ToM scale score (a decimal from 0 to 1) was calculated as a combined proportion calculated as their sum divided by 3. We elected this averaging method because we wanted to give equal weighting to each separate, conceptually distinct aspect of ToM. Cronbach's test revealed adequate internal consistency ($\alpha = .63$) for the Time 1 composite.

Details and rationale for Time 1 EF tests. Two EF tests, route navigation and resisting instructions, exactly replicated Cole and Mitchell's (2000) battery that has been used successfully for groups of a similar age. The former captures EF for forward planning, attention to relevant features, resisting distraction, and the latter, response inhibition. The four route navigation trials were presented and scored (exactly as originally) via maze maps with direct versus indirect paths to a goal and children passed by tracing the former, perfectly, otherwise 0. Sums over trials (0–4) were converted to decimal proportions just as for the combined ToM scores (see above).

Resisting instructions, assessing inhibitory control, was Cole and Mitchell's (2000) adaptation of Reed, Pien, and Rothbart's (1984) test. Questions and procedure were identical to Cole and Mitchell's, but we replaced their squirrel and badger with locally familiar bear and lion puppets. Over 12 trials, the "good" puppet alternated with the "bad" one in issuing commands (e.g., "clap your hands"). Children earned 1 point for each test trial for following the instruction to ignore the latter (i.e., remain still) to a maximum possible task total of 6 points. "Good" puppet (control) trials were not scored. Although there were no practice trials per se, each child had passed a memory control question at the beginning by pointing correctly both at the puppet to be obeyed and the one to be ignored. In line with Cole and Mitchell's procedure, a pause after every fourth trial required children to confirm who should be ignored. All children scored perfectly on these controls. A total proportion score was created by dividing totals by 6.

Time 1 EF total score. Scores on route navigation and resisting instructions were significantly correlated, $r(156) = .39$, $p < .001$, even after age and language were partialed out, $r(154) = .32$, $p < .001$. Hence a composite proportion, the "EF scale score" (0–1) was created.

Time 1 language ability. A widely used, norm-referenced receptive vocabulary test, the PPVT (Dunn & Dunn, 1989), was our measure of language ability. This test of receptive vocabulary was scored according to the test manual to yield raw scores (dubbed PPVT herewith) and standardized verbal mental age estimates (dubbed VMA). Although the test is standardized only on a U.S. population, widespread Australian use of this test (particularly the second edition that we selected for this reason) has confirmed its local suitability (e.g., Oddy et al., 2003).

Time 2 Tasks and Scoring

Time 2 ToM tests. At Time 2 we employed a battery of four ToM tasks that were selected to provide a greater challenge than the tests used at Time 1, thereby making them more appropriate for use with an age group not only 1 year older but also including mostly 5- and early 6-year-olds who normatively pass first-order false belief (Wellman et al., 2001). For continuity with Time 1, the unseen displacement false belief task (two trials) and appearance–reality (two trials) were retained. Then, to increase difficulty and guard against ceiling effects, the remaining two tasks in the Time 2 battery were two more challenging ToM measures requiring emotional inferences. First, we used a real–apparent emotion task closely modeled on one devised by Wellman and Liu (2004). Using Guttman scaling, these authors found that the ToM concept assessed by this task—understanding that someone may display a false emotion on purpose (as when a boy hides his sadness to avoid being called a baby)—is a more advanced ToM concept than false belief. In our version, with exactly Wellman and Liu's pictures, procedure, and question wording, plus an additional stringent control question (as per Peterson, Wellman, & Liu, 2005), children heard a story about a boy, teased by his friends, who did not want them to tease him more. An initial control question asked how the boy really felt. The test question then asked how he made his face look. Correct answers to the latter required both accurate (pictorial or verbal) emotion selection (i.e., one emotion more positive than for the "true feeling") and, as in Peterson et al. (2005), accuracy on a further control question, "Why did he try to look like that?" Children had to display clear comprehension of the protagonist's intention to conceal true feelings to pass the latter. One point was assigned when children answered all questions correctly, otherwise 0.

The second advanced Time 2 ToM task was an emotion false belief test that other authors have shown to be more difficult than first-order false belief (e.g., Bender, Pons, Harris, & de Rosnay, 2011; Hughes et al., 2000). In the version we used, borrowed from Hughes et al. (2000), a puppet who liked animal cookies left and the child helped fill an animal cookie box with rocks. On the puppet's return there were two test questions: (a) "How will [puppet] feel when we give him this all closed up like this?" (happy), (b) "How will he feel when he opens the box and sees rocks?" (sad or angry) plus a control question, and (c) "What does [puppet] like to eat?" (animal cookies). Raw scores (test question sum) were converted to mean proportions (0–1). As noted above, the two-trial unseen displacement false belief test was also retained from Time 1 with new objects, dolls, and hiding places, and scored exactly as at Time 1. The same was true of the two-trial appearance–reality task from Time 1. New objects were a carrot-pen and a dice-shaped jewelry box.

Time 2 ToM total score. All Time 2 ToM measures showed significant intercorrelations (all *ps* $\leq$ .01) that remained uniformly significant with age and PPVT controlled (all *ps* $\leq$ .01). As at Time 1, we computed a total ToM scale score using two steps. First, proportion scores were computed for the unseen displacement false belief test, the emotion false belief test, and the appearance–reality test. These three proportions (0–1) were summed with the real–apparent emotion score (0–1) and divided by 4 to yield a proportion score (0–1). This composite had a Cronbach's alpha of .79, indicating adequate internal consistency.

Time 2 EF tests. At Time 2, four tasks were used to measure EF. We retained the resisting instructions task from Time 1 as it measured inhibitory control, an EF concept of known relevance to preschool ToM (Carlson et al., 2002). Furthermore, this task had proved highly challenging to our sample at Time 1, with few children scoring perfectly (see Table 1). Administration was identical to Time 1, except a bird puppet replaced the "nice" bear, and a monkey, the "naughty" lion. Children's totals (0–6) were converted to proportions (0–1) exactly as at Time 1.

The results for Time 1 (see Table 1) suggested that the route navigation task, already mastered by half the sample at this younger age, might prove insufficiently challenging 1 year later at Time 2. Therefore, we substituted the dimensional card sort task as developed by Frye, Zelazo, and Palfai (1995), a task seen to demand a range of EF capacities including resistance to distraction, response inhibition, memory, and attention to relevant features. As in Frye et al., children were provided with a pile of cards, each depicting either a yellow circle or a red square. Behind each of two boxes stood an "identification" card: either a red circle or a yellow square. Children were initially trained to sort the cards on the basis of one dimension (e.g., shape). Then, after a distraction, they were trained on a reversal shift requiring sorting by a different dimension (e.g., color). Finally, after another distraction there were three test trials. The first required children to sort five cards according to the most recent training dimension. One point was awarded for each correct placement, to a total of 5. Next, the rule was reversed (without further training) and children sorted five more cards, earning 1 point each if correct. Finally, four trials were given with the sorting rule alternating between every trial. Again each correct answer scored one point. Total possible scores, combining the three tests, thus ranged from 0 to 14 and were divided by 14 to give a proportion score, in line with other tasks.

Another Time 2 EF test, the hand game, was included to assess inhibitory control, an EF dimension shown previously to have close links with ToM (Carlson et al., 2002). Children were told: "We are going to play a game. First we both put our hands behind our backs; now when I show my hand I want you to make the same shape as me. So if I make a fist, you make a fist, and if I point a finger you point a finger." After copying (control condition) the experimenter's actions to a total of six consecutive correct responses (out of a maximum of 15 trials which no child in this sample required) the test, a conflict condition, was given. The tester said, "Now, if I point a finger, I want you to show a fist. And if I show a fist I want you to point a finger. So we're not making the same shapes." Children earned higher scores the fewer trials they needed to reach the criterion of six consecutive correct responses. Any child not achieving criterion within 15 test trials scored 0, whereas 1 point was given for requiring the full 15 trials, 2 points for requiring only 14, 3 for 13, and so forth. The maximum possible score was 10 if the child's performance was errorless. Raw scores (see Table 1) were divided by 10 to convert to proportion scores ranging from 0 to 1.

The Tower of London EF task, adapted from Shallice (1982), was chosen as a challenging measure of strategic forward planning, attention, and working memory. Ours was adapted for younger children by using only three pegs (of three different heights). Three hollow colored balls (red, blue, green) could slide onto the pegs: The shortest peg could hold only one; the tallest, all three. Problems involving movement of balls were presented using three-dimensional models. A warm-up set, presented first, involved need for only a single move and was not scored. On the fourth and subsequent trials (comprising successively three of each of two-move trials, three-move trials, and four-move problems), children scored 2 on each trial for succeeding with minimum moves. One move above minimum scored 1, otherwise 0. After three consecutive 0s, the task was terminated. Scores per trial (0–2) were summed and the total (0 to 24) was divided by 24 to give a proportion score as for other tasks.

Time 2 EF total score. All Time 2 EF measures showed significant correlations with one another (all $rs \geq .23$, all $ps \leq .01$). These correlations remained significant after controlling for age and language (all $rs \geq .21$, all $ps \leq .05$). Cronbach's alpha for the four-task composite was .86, indicating good internal consistency. Time 2 EF total scale scores were again computed proportionally to reflect a decimal from 0 to 1.

Time 2 language ability. The PPVT was again administered in full at Time 2, and raw PPVT scores were again computed and used in all statistical analyses.

Results

Overview

Table 2 shows the raw correlations among all key variables. Partial correlations (controlling for age and raw PPVT scores) also appear in this table. Correlations across the testing times (i.e., from Time 1, the initial test, to Time 2, 1 year later) are shown. We explore these in relation to two main research questions: (a) How does having child-aged siblings relate to ToM and EF performance at Times 1 and 2? and (b) What do cross-lagged correlations say about how gains over time in EF and ToM capacities relate to earlier individual differences in these variables independently of age, language ability, and number of child siblings? (Each of the analyses outlined below was run again with the exact lag from Time 1 to Time 2 entered as a further covariate. As all findings remained unchanged, results are not reported here.)

Siblings and the Development of ToM and EF Skills

Table 2 shows the correlations between total child-aged siblings with total ToM and EF proportion scores at Times 1 and 2. In addition, to more fully examine the role of siblings in children's performance on ToM and EF tasks at both time points, the information regarding number of child-aged siblings was recoded to yield an additional two different sibling variables. Thus, in addition to examining the influence of the number of child-aged siblings, we also examine presence versus absence of child-aged siblings (dummy coded as 1 or 0), as well as the child's birth-order position in the sibling constellation (1 = *youngest in family*, 2 = *middle child*, 3 = *eldest in family*). Correlations of these variables with ToM and EF at each testing time also appear in Table 2.

The correlations suggest that the presence of a greater number of child siblings contributes to a

Table 2

Raw and Partial (Controlling for Age and Language Ability) Correlations Between all Key Variables at Times 1 and 2

	Peabody Picture Vocabulary Test (PPVT)	Time 1 variables							Time 2 variables				
		No. sibs	Sibs present	Sibling situation	ToM	EF	Age	PPVT	No. sibs	Sibs present	Sibling situation	ToM	EF
T1 age[a]	.62***	.34*** [.30***]	.25** [.23**]	.22** [.19*]	.33*** [.17*]	.52*** [.43***]	.94*** [.91***]	.60*** [−.17]	.35*** [.33***]	.15 [.17*]	.08 [.11]	.13 [.08]	.18* [.07]
T1 PPVT[b]	—	.18* [−.04]	.13 [−.05]	.12 [−.03]	.32*** [.16]	.32*** [−.00]	.50*** [−.02]	.82*** [.84***]	.14 [−.10]	.02 [−.09]	−.01 [−.07]	.13 [.04]	.20* [.12]
T1 number of siblings[c]		—	.60*** [.62***]	.32*** [.26**]	.38*** [.32***]	.30*** [.20*]	.60*** [.58***]	.10 [.09]	.82*** [.89***]	.58*** [.57***]	.19* [.17*]	.30*** [.26**]	.26** [.22**]
T1 presence or absence of siblings[c]			—	.59*** [.57***]	.23** [.19*]	.32*** [.23**]	.37*** [.28**]	.18* [.17*]	.54*** [.55***]	.78*** [.87***]	.37*** [.36***]	.19* [.07]	.25** [.19*]
T1 sibling situation[c]				—	.08 [.02]	.26** [.17*]	.16 [.14]	.05 [.02]	.26** [.20*]	.48*** [.47***]	.82*** [.83***]	−.26** [−.30***]	.08 [.04]
T1 ToM scale score[c]					—	.28*** [.14]	.35*** [.32***]	.42*** [.38***]	.35*** [.28***]	.17* [.14]	−.03 [−.05]	.24** [.20*]	.32*** [.27***]
T1 EF scale score[c]						—	.60*** [.57***]	.41*** [.38***]	.27** [.11]	.22** [.17*]	.11 [.09]	.03 [−.04]	.26** [.20*]
T2 age[d]							—	.50*** [.42***]	.52*** [.38***]	.33** [.30***]	−.00 [.18*]	.32*** [−.01]	.18* [.09]
T2 PPVT[e]								—	.15 [−.01]	.08 [−.13]	−.09 [−.29**]	.36*** [.23**]	.20* [−.07]
T2 number of siblings[f]									—	.64*** [.46***]	.27** [.13]	.29*** [.23**]	.23** [.21*]
T2 presence or absence of siblings[f]										—	.56*** [.46**]	.12 [−.19]	.14 [−.01]
T2 sibling situation[f]											—	−.23** [−.51**]	−.04 [−.04]
T2 ToM scale score[f]												—	.14 [.01]
T2 EF scale score[f]													—

[a] T1 PPVT controlled. [b] T1 age controlled. [c] T1 age and PPVT controlled. [d] T2 PPVT controlled. [e] T2 age controlled. [f] T2 age and PPVT controlled.
*$p < .05$. **$p < .01$. ***$p < .001$.

preschool child's ToM and EF development. Even after partialing out influences from age and raw PPVT scores, number of child siblings is significantly related to ToM at Time 1, $r(154) = .32$, $p < .001$, and Time 2, $r(154) = .26$, $p < .01$, as well as EF at Time 1, $r(154) = .20$, $p < .05$, and Time 2, $r(154) = .22$, $p < .01$. Here, we question whether ToM and EF are enhanced simply as a linear function of number of child siblings or if the relations observed thus far might simply reflect an advantage of sibling presence (in any number) over the absence of siblings. We start with an examination of sibling presence versus absence.

Sibling presence versus absence as it relates to ToM and EF. A one-way analysis of covariance (ANCOVA) (with age and PPVT raw scores as controls) was used to examine the mere presence of sibling(s) (0 = *only-child*) as a predictor of Time 1 ToM. A similar ANCOVA was then run with Time 2 ToM as the dependent variable. For Time 1 ToM, the results were statistically significant in the anticipated direction, with only-children ($n = 24$) scoring lower on ToM ($M = .43$, $SD = .24$) than children with at least one sibling ($n = 133$, $M = .61$, $SD = .25$), $F(1, 153) = 5.76$, $p < .05$, partial $\eta^2 = .04$. In other words, simply having a child sibling at all was of benefit. Similar results were observed in the ANCOVA using Time 2 ToM as the dependent variable. Here, only-children ($n = 21$) with a mean ToM score of .45 ($SD = .27$) were found to score significantly lower, $F(1, 153) = 4.10$, $p < .05$, partial $\eta^2 = .03$, than children with at least one sibling ($n = 136$, $M = .60$, $SD = .28$). The results of these ANCOVA analyses consistently support the notion of a disadvantage, at both testing times, of only-child status for ToM mastery.

Similar ANCOVAs were used to examine EF as a function of sibling presence or absence after controlling for age and PPVT scores. With Time 1 EF as the dependent variable, the results paralleled those for ToM. Only-children ($n = 24$) earned significantly lower EF scores at Time 1 ($M = .51$, $SD = .18$) than children with at least one sibling ($n = 133$, $M = .75$, $SD = .25$), $F(1, 153) = 8.04$, $p < .01$, partial $\eta^2 = .05$. A similarly significant effect was observed for EF scores at Time 2. Only-children ($n = 21$) with a mean of .69 ($SD = .22$) were outperformed by children with at least one sibling ($n = 136$, $M = .81$, $SD = .15$), $F(1, 153) = 7.67$, $p < .01$, partial $\eta^2 = .05$. After taking account of age and language ability, only-child status appeared to be a disadvantage for EF development for children at both the age periods, just as it was for ToM.

Number of child-aged siblings as it relates to ToM and EF. To test whether larger numbers of child-aged siblings were linked to higher Time 1 or Time 2 ToM or EF scores, we ran a similar set of ANCOVAs controlling for age and PPVT. The sibling variable was recategorized as "1 child sibling," "2 child siblings," or "3 or more child siblings" to avoid distortion by the few participants with more than three. To avoid duplicating the previous set of regressions, we used only the children with at least one sibling in these analyses. (Note, however, that Table 2 includes correlations from the full sample, including only-children.)

Number of child-aged siblings was a significant overall predictor of Time 1 ToM, $F(2, 128) = 7.21$, $p < .01$, partial $\eta^2 = .10$, supporting our hypothesis. Planned comparisons revealed no significant difference between children with one versus two siblings ($n = 74$, $M = .56$, $SD = .25$ and $n = 41$, $M = .60$, $SD = .20$, respectively), contrast estimate = .03, *ns*. However, children with three siblings or more ($n = 18$, $M = .83$, $SD = .25$) outperformed those with two siblings (contrast estimate = .23, $p < .01$), as well as those with only one sibling (contrast estimate = .32, $p < .001$). Time 2 ToM revealed the same pattern. Number of child siblings was the only significant variable in the model, $F(2, 131) = 6.89$, $p < .01$, partial $\eta^2 = .10$. At Time 2, there was no significant difference between those with one sibling ($n = 66$, $M = .53$, $SD = .25$) and those with two siblings ($n = 48$, $M = .59$, $SD = .23$), contrast estimate = .06, *ns*. However, children with three or more siblings at Time 2 scored higher on Time 2 ToM ($n = 22$, $M = .79$, $SD = .18$) than those with two (contrast estimate = .20, $p < .01$) or one (contrast estimate = .26, $p < .001$). These findings consistently support the hypothesized advantage, at both testing times, of access to more siblings for rapid ToM mastery.

Similar ANCOVAs were conducted to examine the association between EF and number of child siblings at each time point. In the analysis of Time 1 EF, the model was significant, $F(4, 128) = 10.31$, $p < .001$, partial $\eta^2 = .24$. However, age was the only significant contributor, $F(1, 128) = 25.11$, $p < .001$, partial $\eta^2 = .16$. Number of child siblings was not a significant predictor, $F(2, 128) = .91$, *ns*. In the analysis of Time 2 EF, the model was once again significant, $F(4, 131) = 40.42$, $p < .01$, partial $\eta^2 = .26$. Number of child siblings was significant, $F(2, 131) = 4.33$, $p < .05$, partial $\eta^2 = .06$. Contrast estimates revealed that children with two siblings ($n = 48$, $M = .85$, $SD = .11$) outperformed children with one sibling ($n = 66$, $M = .76$, $SD = 17$), contrast estimate = .09, $p < .01$. No other contrasts were significant.

Birth-order position in the sibling constellation as it relates to ToM and EF. Next, we sought to determine whether a child's position in the family's sibling constellation (eldest, youngest, or middle child) is important. To answer this question, data from only-children were set aside and additional ANCOVAs were conducted on just those children who had at least one sibling. Dubbed "sibling constellation," we created three categories: those who were (a) youngest in the family ($n = 97$), (b) middle children ($n = 13$), or (c) eldest of siblings ($n = 23$).

First, in an ANCOVA with Time 1 ToM as dependent variable, we controlled for age, PPVT raw scores, and total sibling number (given that middle status implied at least two siblings, whereas many with only older or younger siblings, as for $n = 74$ in our Time 1 sample, had just one). In relation to T1 ToM, the overall model was statistically significant, $F(5, 127) = 7.98$, $p < .001$, partial $\eta^2 = .24$. The age covariate was not significant. However, the covariate of PPVT raw scores was significant, $F(1, 127) = 5.82$, $p < .05$, partial $\eta^2 = .04$, as was the covariate of total siblings, $F(1, 127) = 7.48$, $p < .01$, partial $\eta^2 = .06$. Moreover, the target variable (sibling constellation) was not significant once these factors were statistically controlled, $F(2, 127) = 1.48$, *ns*. In other words, just having younger (child-aged) siblings was no less advantageous than having older siblings or being a middle child for Time 1 ToM in this sample.

An ANCOVA with the same design was then run with Time 2 ToM as the dependent variable. At Time 2 there were 97 youngest children, 13 middle, and 26 eldest in the family. Sixty-six eldest and youngest children had only one child-aged sibling. In relation to Time 2 ToM, the model was significant, $F(5, 130) = 10.82$, $p < .001$, partial $\eta^2 = .32$. Neither age nor PPVT contributed significantly ($ps = .28$ and $.90$, respectively). Number of child siblings was a statistically significant covariate, $F(1, 130) = 13.82$, $p < .001$, partial $\eta^2 = .10$, and (as predicted, but in contrast to Time 1 ToM) the focal sibling constellation variable was also statistically significant, $F(2, 130) = 18.10$, $p < .001$, partial $\eta^2 = .22$. The sibling advantage—in terms of ToM development in late preschool and early primary school—appears to accrue most strongly for children with older siblings (i.e., those who are the youngest in the family). Planned contrasts showed that the estimated marginal mean of Time 2 ToM scores for youngest children ($M = .68$) was greater than that for middle children ($M = .42$, contrast estimate $= -.27$, $p < .001$), and also eldest children ($M = .41$, contrast estimate $= -.25$, $p < .001$).

We conducted similar ANCOVAs using EF scores as the dependent variable, first at Time 1 and then at Time 2. Birth-order position (eldest, etc.) was not significantly related to EF at either time point (both $Fs < 1.67$, both $ps > .18$). For Time 1 EF the overall model was significant, $F(5, 127) = 7.94$, $p < .001$, partial $\eta^2 = .24$, but the age covariate was the only significant predictor, $F(1, 127) = 22.22$, $p < .001$, partial $\eta^2 = .15$. For Time 2 EF, the model was again significant, $F(5, 130) = 2.70$, $p < .05$, partial $\eta^2 = .11$, but the pattern was different from that at Time 1. Total child siblings were the only significant Time 2 predictor, $F(1, 130) = 4.22$, $p < .05$, partial $\eta^2 = .03$.

Summary of sibling associations with ToM and EF. From the analyses thus far, we have learned that influences from child siblings are relevant to ToM and EF at both time points. Having any child-aged sibling at all (i.e., not being an only-child) appears to be consistently related to advanced ToM and EF. Within the subgroup with at least one sibling, ToM was more advanced for those with the largest family size. EF was less consistently influenced, although the pattern of results at Time 2 was similar to that for ToM. Once the variables of age, language, and total family size were statistically controlled, however, there was less consistent evidence for any distinct influence of particular birth-order positions on either ToM or EF. In other words, sibling presence (vs. absence) and total family size each were significantly related to ToM and EF at Times 1 and 2. Birth order in the sibling constellation only added significantly in one case. At Time 2, being the youngest sibling was associated with more advanced ToM over and above the effects associated with having more siblings, being older, and having more advanced language.

Predictive Relations Among EF, ToM, and Siblings Over Time

In addition to examining sibling effects on ToM and EF performance at Times 1 and 2, another aim of this study was to examine relations between individual differences in ToM and EF across the age range 3 years 3 months to 6 years 8 months, using cross-lagged longitudinal correlations. While not unequivocally identifying cause–effect relations among variables, cross-lagged correlations (in which performance on the target variable at the initial test time is statistically controlled to examine later progress in that same variable) are useful for supplying novel information on the place of EF skills as developmental precursors versus chronological sequels to gains in ToM understanding, and vice versa. As well as controlling for the earlier levels of the outcome variable via the cross-lagged method, we also controlled statistically for the effects of age, verbal ability, and number of child siblings.

Two separate hierarchical regression analyses were conducted to assess these effects. This was because we had no a priori assumptions, or past findings to rely upon, for the present age group. In each hierarchical regression, Time 1 age and Time 1 PPVT scores were entered in the first block, as control variables, along with the Time 1 score on the target-dependent variable (i.e., Time 1 EF was partialed out of the final regression when Time 2 EF was the outcome variable, and Time 1 ToM was controlled when Time 2 ToM served as the outcome). The second block was used to assess the impact of the independent variable of interest (e.g., effect of Time 1 ToM on Time 2 EF). Finally, number of child siblings at Time 1 was entered in the third block in each equation to see if this variable made any independent contribution over and above all other variables already in each model.

Time 1 ToM as a predictor of Time 2 EF. The hypothesis that early ToM understanding predicts later advances in EF test performance was tested with a hierarchical multiple regression that had Time 2 EF as the dependent variable. When Time 1 age, PPVT scores, and EF scores were entered as controls in the first block, the regression equation was statistically significant, $R = .29$, $R^2 = .08$, $F_{\text{change}}(3, 153) = 4.64$, $p < .01$. As shown in Table 3, the beta weight for Time 1 EF was significant in this first block, but those for age and language were not.

Time 1 ToM was entered in the second block. The result was a significant increment, $R = .38$, $R^2 = .14$, $\Delta R^2 = .06$, $F_{\text{change}}(1, 152) = 10.08$, $p < .01$, indicating Time 1 ToM made a unique contribution to the Time 2 EF variability that was predicted. The inclusion of number of child siblings at the final step did not result in any increase in explained variance, $R = .40$, $R^2 = .16$, $\Delta R^2 = .02$, $F_{\text{change}}(1, 151) = 2.99$, *ns*. Results for the full model at the end of this third block are displayed in Table 3. Time 1 ToM was a statistically significant predictor of Time 2 EF, $t(151) = 2.61$, $p < .05$, and was the only significant predictor variable in the final equation. In other words, early understanding of minds was linked with later gains in EF skills independently of the other variables. These results are consistent with the developmental precedence of ToM-based understanding of mental states (like false beliefs) as a link to later abilities to deploy cognitive EF skills (like inhibitory control or strategic memory) more effectively. Note that this finding for the age period from 4 to 5 years that we studied contrasts with a predominantly opposite pattern in past studies of younger age groups, as we discuss in more detail later.

Time 1 EF as a cross-lagged predictor of Time 2 ToM. We conducted a further hierarchical regression to test whether Time 2 ToM might be influenced by Time 1 EF. The model was set up as above, except that ToM variables were substituted for EF variables, and vice versa. Results appear in Table 4.

Table 3
Hierarchical Regression Predicting Time 2 EF From Time 1 ToM, While Controlling for Age, Language, Time 1 EF, and Number of Child Siblings

Block	Variable	β	sr_i	sr_i^2	t
1	Age	−.03	−.02	.00	−0.27
	Peabody Picture Vocabulary Test (PPVT)	.14	.11	.01	1.47
	Time 1 EF	.23	.20	.04	2.55*
2	Age	−.07	−.05	.00	−0.65
	PPVT	.09	.07	.00	0.89
	Time 1 EF	.18	.15	.02	1.97*
	Time 1 ToM	.27	.24	.06	3.17**
3	Age	−.11	−.07	.00	−0.97
	PPVT	.10	.08	.01	1.05
	Time 1 EF	.16	.13	.02	1.79
	Time 1 ToM	.23	.20	.04	2.61*
	Number of child siblings	.15	.13	.02	1.73

Note. EF = executive functioning; ToM = theory of mind.
$^*p < .05$. $^{**}p < .01$.

Table 4
Hierarchical Regression Predicting Time 2 ToM From Time 1 EF, While Controlling for Age, Language, Time 1 ToM, and Number of Child Siblings

Block	Variable	β	sr_i	sr_i^2	t
1	Age	.04	.03	.00	0.43
	Peabody Picture Vocabulary Test (PPVT)	.01	.01	.00	0.09
	Time 1 ToM	.21	.19	.04	2.43*
2	Age	.09	.06	.00	0.78
	PPVT	.01	.00	.00	0.05
	Time 1 ToM	.23	.20	.04	2.57*
	Time 1 EF	−.09	−.08	.01	−0.99
3	Age	.03	.02	.00	0.23
	PPVT	.03	.03	.00	0.33
	Time 1 ToM	.16	.13	.02	1.74
	Time 1 EF	−.12	−.10	.01	−1.30
	Number of child siblings	.25	.23	.05	2.93**

Note. EF = executive functioning; ToM = theory of mind.
$^*p < .05$. $^{**}p < .01$.

With control variables of age, language, and Time 1 ToM entered, the first block was significant, $R = .23$, $R^2 = .06$, $F_{\text{change}}(3, 153) = 2.97$, $p < .05$. Neither age nor PPVT raw scores emerged as significant, $ts(153) = 0.43$ and 0.09, respectively, both $ps > .66$, but Time 1 ToM significantly predicted Time 2 ToM, $t(153) = 2.43$, $p < .05$. Adding Time 1 EF in the second block failed to increase explained variance, $R = .25$, $R^2 = .06$, $\Delta R^2 = .00$, $F_{\text{change}}(1, 152) = 0.97$, *ns*. However, when the number of child siblings was entered in the third block, there was a significant change, $R = .33$, $R^2 = .11$, $\Delta R^2 = .05$, $F_{\text{change}}(1, 151) = 8.60$, $p < .01$. In the final model, the influence of number of child siblings eclipsed all other variables, including Time 1 EF, with the only significant beta at the final step (see Table 4). These results suggest that Time 2 ToM scores were not significantly influenced by Time 1 EF over the age range that we examined. Interestingly, though, number of child siblings was a significant predictor of 5-year-old ToM performance, above all other variables.

Discussion

These data contribute to a well-established literature linking ToM and EF developments among preschoolers. The results are novel, however, in several key respects. First, we used a longitudinal cross-lagged correlation design with a somewhat older age group than in most past research. We found, for the first time, that ToM performance at a mean age of 4.17 years was significantly related to individual differences in children's EF scores 1 year later ($M_{\text{age}} = 5.17$). Furthermore, these Time 1 ToM effects were significant over and above the effects of Time 1 EF upon EF skills at Time 2, and emerged from analyses that also controlled statistically for effects of age, language skill, and number of child siblings. In other words, how advanced a child was in earlier mental state understanding independently predicted EF skills at age 5, even after taking account of other known correlates including language and Time 1 EF abilities.

This result was unpredicted and evidences a pattern that appears relatively novel in the published literature. As noted earlier, cross-sectional links between ToM and EF skills among preschool children are well documented (Moses et al., 2006). Fewer studies have used the more informative cross-lagged longitudinal approach and these have focused mainly on children who were substantially younger than those in the present sample. While yielding valuable insight into the earliest associations of ToM and EF in toddlerhood and preschool, little published data track these relations longitudinally further in development (but see Farrant et al., 2012, for an exception). Importantly, when this was done in this study, the pattern differed from that most often observed previously. Rather than early EF statistically predicting later ToM as is frequently seen in younger groups (e.g., Carlson et al., 2004; Hughes & Ensor, 2007; Müller et al., 2012), our findings for an age period spanning, for most children, the transition from a play-oriented preschool curriculum to a more formal academic one in upper pre- or early primary school, displayed the opposite pattern. Earlier ToM emerged as a cross-lagged EF predictor over and above other Time 1 and Time 2 correlates. Our results show this prerequisite more clearly and less equivocally than any previous study (including that of Hughes and Ensor, 2007, who found no cross-lagged prediction of later EF from earlier ToM between ages 2 and 3 or ages 3 and 4, but some suggestion of longer term prediction from ages 2 to 4). Possibly the pathway emerged more clearly in our study because our sample was somewhat older. The more adept EF deployment by older preschoolers or young school-aged children who had been precocious in ToM understanding 1 year earlier appears consistent with Perner and Lang's (1999) general suggestion that children's EF development may be enhanced by social cognitive understanding of minds. Perhaps this is particularly true of the understanding of false belief, a capacity that does not normatively emerge consistently until after ages 4.5 or 5 years (Wellman et al., 2001). Cole and Mitchell's (2000) suggestion that social and conversational interaction with siblings (particularly in contexts of disputes and rule games) might, in theory, benefit both EF and ToM is also broadly supported by the present findings.

Nevertheless, there are many other possible accounts for these results, and further investigation and replication is clearly warranted. It would be particularly useful for further studies to examine the sibling variable more fully, including direct observation of interactions among siblings of different ages in more naturalistic ToM-relevant and EF-relevant everyday situations (e.g., while playing card games, board games, or deceptive games like hide-and-seek). Also, as Sabbagh, Xu, Carlson, Moses, and Lee (2006) noted, it would be valuable to include different cultural groups besides Anglo-heritage speakers of English in these studies. Indeed, in cross-cultural comparison of Chinese and U.S. preschoolers, Sabbagh et al. found that Chinese children (none of whom had siblings) significantly outperformed their U.S. counterparts (virtually all of whom had at least one sibling) on EF tasks. However, the Chinese group was no more advanced in ToM performance despite this early EF advantage. Although many cultural or sampling factors besides (or in addition to) siblings could potentially account for these results, Sabbagh et al. noted the absence of sibling–sibling interaction owing to China's one child family policy was a plausible candidate. Indeed, those in our Australian sample who were only-children were outperformed on both ToM and EF at both time points.

There are also valid theoretical reasons to speculate that having child siblings at home with whom to play and converse may be advantageous to early social-cognition and early EF performance. Make-believe play, disputes, reminiscence, and other social and conversational interactions among young siblings offer rich and varied exposure to others' minds in ways likely to be useful for early ToM understanding (Harris, 2005). At the same time, in relation to EF, having a child sibling boosts opportunities to compete in games and serious pursuits with an equal, and this potentially challenges children to develop planning, memory, strategic thought, and inhibitory control (Cole & Mitchell, 2000). Such uniquely sibling-based experiences could, in theory, accelerate both ToM and EF growth.

Indeed, a large body of past research links access to child-aged siblings with rapid ToM growth (e.g., Jenkins & Astington, 1996; McAlister & Peterson, 2007; Ruffman et al., 1998) and EF (e.g., Hughes & Ensor, 2005). The sibling link to EF has less often been confirmed (Carlson et al., 2004; Cole & Mitchell, 2000), though, in one previous study that took account of the sibling's age (a variable likely to place limits upon the opportunity to profit from sibling interaction; Peterson, 2000), McAlister and Peterson (2006) found cross-sectionally that number of child-aged siblings was a significant raw correlate of preschoolers' EF scores. However, they also found that ToM, but not EF, exerted a unique contribution when age and language ability were statistically controlled and suggested that number of child siblings might not contribute to EF independently of ToM.

This conclusion, although provocative, was limited in the past by the cross-sectional nature of the data used to support it. The present findings, via the use of a longitudinal design, both replicated and extended McAlister and Peterson's (2006) results. When we included number of child siblings as a potential explanatory variable, our results supplied important, novel information. The longitudinal results showed that children with siblings at Time 1 scored higher on both EF and ToM, not only contemporaneously but also longitudinally 1 year later. Furthermore, these effects were independent of both age and language skill. We also explored dimensions of the sibling variable in more detail and found that number of child siblings at Time 1 remained a significant predictor of Time 2 ToM understanding over and above not only chronological and linguistic maturity but also Time 1 levels of both ToM and EF skills.

In other words, at a mean age of 5 years, the children in our sample who had had opportunities to interact with child siblings at home as younger preschoolers were better at understanding others' minds irrespective of any benefits accruing from age, advanced language or EF skills. It is important that this sibling–ToM association persisted at Time 2 when many children had entered full-day primary schooling. Furthermore, when we closely examined the sibling variable in terms of its separate aspects, we found that having a child sibling at all (as opposed to being an only-child) was associated with more advanced ToM and EF performance at both Time 1 and Time 2 and that this was true over and above the potential confounds of age and language ability. However, presence or absence was not the only significant sibling effect. When we excluded only-children greater total numbers of child-aged siblings predicted advanced ToM at Times 1 and 2, over and above other variables. Furthermore, the benefit accrued to children who were eldest (and thus had no older siblings) as much as to those with siblings in other birth-order positions. For EF, by contrast, the primary predictor was the fact of not being an only-child. These results for ToM and EF make sense theoretically. Siblings' distinctive patterns of conversation, play, and social interaction are likely to supply insight into others' thoughts and feelings (Harris, 2005) and, when unavailable, any ToM advantages conferred by early mastery of EF skills may be masked by the lack of exposure to these early sibling-based sources of ToM stimulation, as has been posited for children growing up in China (Sabbagh et al., 2006).

Interestingly, however, our data suggest that siblings' beneficial effects are direct for ToM, whereas for EF they seem to be channeled at least partially through early ToM development. Young children with extensive sibling experience are believed to achieve ToM understanding precociously via their rich exposure to conversations, play, disputes, and other interactions (Harris, 2005). Such early expertise in "reading" minds would then plausibly lead to further rich and varied social participation with siblings and peers in varied peer-based social activities (Peterson, Slaughter, & Paynter, 2007) like pretend play, strategic deception, and games. Thus, a young 4-year-old preschooler with advanced ToM understanding and rich sibling play experience would be well equipped to take further advantage of the kinds of play experiences that might most rapidly boost EF skills.

The process is undoubtedly a mutual one, however, with advanced ToM assisting gains in EF (as well as further ToM gains) and vice versa. For example, to be a valued partner in games, preschoolers must inhibit impulses to cheat, grab, speak out of turn, or otherwise undermine the game. Much childhood play requires EF inhibition skills (e.g., to avoid calling out in hide-and-seek or treating the pretense as real). Equally, ToM-driven understanding of the motivations and mental states of other players is crucial to such play. Furthermore, at a metalevel, abilities to recognize the play partner's desires (e.g., for a cooperative, rule-abiding playmate), together with strategic and mindful negotiation and conflict-resolution skills are fundamental to most forms of satisfying social interactions. Thus, ToM and EF together may both build upon past social experience (including sibling interaction) and facilitate further social participation.

Of course, it is important for these findings to be replicated before any interpretations can be more than purely speculative. Further longitudinal research with both older and younger samples would also be useful to more fully explore possible changes in the nature of the association between EF and ToM with development, in line with present findings. As Best et al. (2009) noted, EF skills develop importantly from age 5 onward so that samples of children in most past cross-lagged research may have been too young to optimally reveal how ToM contributes to the later growth of these later emerging EF capacities. Furthermore, our evidence as to the importance of sibling influences may also be somewhat age dependent. The young children studied in most past longitudinal EF–ToM research may not yet have been as capable as those in our somewhat older sample of gaining ToM or EF benefits from conversational and social interactions with their siblings, particularly with siblings younger than themselves. Moreover, with an older sample of children (who may be more likely to have a greater number of siblings), it may be possible to test whether the sibling effect is curvilinear. It might well be that more siblings is "better," but only up to a point. Because very few of the children in our sample had more than two siblings, and only one had more than three, a curvilinear relation could not effectively be tested with present data. Furthermore, our study was limited by the relatively broad age range of the sample at each testing point. Few measures of EF and ToM optimally suit the entire age range tested. Hence, regardless of the age group of interest, future longitudinal research might do well to sample larger numbers within a narrow age band at each point and, despite known individual variability in both EF and ToM among children of identical age (Dunn, 1995), might help to circumvent the limitations of the present research that some tasks showed ceiling and others floor effects at one of the time points. Owing to the relatively wide age range at each testing time, some of the eldest children in the sample were less than optimally challenged by tests that best suited the youngest participants, and vice versa. Furthermore, it would be helpful in future research to adopt larger test batteries for ToM and especially for EF than in this and much past research. Our study had only two EF tests at Time 1 and four at Time 2. Farrant et al. (2012) had only one at each testing point. A more comprehensive battery would ensure that all EF's multiple facets were encompassed.

Despite these limitations, we believe our findings, although requiring replication and extension, make two important contributions. One is methodological and the other more substantive. Methodologically, in line with Sabbagh et al.'s (2006) suggestion, they highlight the importance of including assessment of participants' sibling status in future cross-sectional and longitudinal studies of the ToM–EF relation. This hitherto largely neglected variable clearly has important explanatory potential that could, as in this study, assist in clarifying developmental patterns. Going beyond the present findings, it will also be useful in future work to explore the specific kinds of interactions among siblings that are most closely linked with advances in ToM and EF skills. In terms of theory, our finding that Time 1 ToM contributes significantly to Time 2 EF, together with several past studies showing early EF's contribution to later ToM, suggest that neither set of skills completely precedes or is uniquely prerequisite to the other across early development. Instead, when the body of longitudinal research as a whole is considered, it would seem that advances in both cognitive domains assist reciprocally one another, especially when the catalyst of early sibling interaction is added to the equation. Future research could profitably explore these exciting possibilities.

References

Australian Bureau of Statistics. (2008, February 29). *2006 Census community profile series: 4069 (postal area).* Retrieved July 20, 2011, from http://www.censusdata.abs.gov.au

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37–46. doi:10.1016/0010-0277(85)90022-8

Bender, P. K., Pons, F., Harris, P. L., & de Rosnay, M. (2011). Do young children misunderstand their own emotions? *European Journal of Developmental Psychology, 8*, 331–348. doi:10.1080/17405629.2010.495615

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions at age 5: Changes and correlates. *Developmental Review, 29*, 180–200. doi:10.1016/j.dr.2009.05.002

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology, 40*, 1105–1122. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053. doi:10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92. doi:10.1002/icd.298

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology, 18*, 279–295. doi:10.1348/026151000165698

Dunn, J. (1995). Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. *Cognition and Emotion, 9*, 187–201. doi:10.1080/02699939508409008

Dunn, J., & Brophy, M. (2005). Communication, relationships, and individual differences in children's understanding of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 50–69). New York: Oxford University Press.

Dunn, L. M., & Dunn, L. M. (1989). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Farrant, B. M., Maybery, M. T., & Fletcher, J. (2012). Language, cognitive flexibility, and explicit false belief understanding: Longitudinal analysis in typical development and specific language impairment. *Child Development, 83*, 223–235. doi:10.1111/j.1467-8624.2011.01681.x

Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist, 41*, 418–425. doi:10.1037/0003-066X.41.4.418

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly, 50*, 274–290.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science, 7*, 103–115. doi:10.1111/j.1467-7687.2004.00326.x

Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. Theory of mind in typical and atypical development. *Current Directions in Psychological Science, 7*, 116–121.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527. doi:10.1016/0885-2014(95)90024-1

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60. doi:10.1037/0033-2909.134.1.31

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*, 26–37.

Harris, P. L. (2005). Conversation, pretense, and theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 70–83). New York: Oxford University Press.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253. doi:10.1111/j.2044-835X.1998.tb00921.x

Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry, 41*, 483–490.

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology, 28*, 645–668. doi:10.1207/s15326942dn2802_5

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*, 1447–1459. doi:10.1037/0012-1649.43.6.1447

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32*, 70–78. doi:10.1037/0012-1649.32.1.70

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251. doi:10.1016/0010-0277(92)90013-8

McAlister, A. R., & Peterson, C. (2006). Mental playmates: Siblings, executive functioning and theory of mind. *British Journal of Developmental Psychology, 24*, 733–751. doi:10.1348/026151005X70094

McAlister, A. R., & Peterson, C. C. (2007). A longitudinal study of child siblings and theory of mind development. *Cognitive Development, 22*, 258–270. doi:10.1016/j.cogdev.2006.10.009

Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2006). On the specificity of the relation between executive functioning and children's theories of mind. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 131–146). Mahwah, NJ: Erlbaum.

Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I. M., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of Experimental Child Psychology, 111*, 331–348. doi:10.1016/j.jecp. 2011.08.014

Oddy, W., Kendall, G., Blair, E., de Klerk, N., Stanley, F., Landau, L., et al. (2003). Breast feeding and cognitive development in childhood. *Paediatric & Perinatal Epidemiology, 17*, 81–90. doi:10.1046/j.1365-3016.2003.00464.x

Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development, 79*, 80–99. doi:10.1111/j.1467-8624.2007.01112.x

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*, 1081–1105. doi:10.1111/j.1469-7610.1991.tb00351.x

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Perner, J. (1998). The meta-intentional nature of executive functions and theory of mind. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 270–283). Cambridge, England: Cambridge University Press.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*, 337–344. doi:10.1016/S1364-6613(99)01362-5

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73*, 752–767. doi:10.1111/1467-8624.00436

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*, 1228–1238. doi:10.1111/j.1467-8624.1994.tb00814.x

Peterson, C. C. (2000). Kindred spirits: Influences of siblings' perspectives on theory of mind. *Cognitive Development, 15*, 435–455. doi:10.1016/S0885-2014(01)00040-5

Peterson, C. C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf, and normal hearing children. *Psychological Science, 10*, 126–129. doi:10.1111/1467-9280.00119

Peterson, C. C., Slaughter, V. P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry, 48*, 1243–1250. doi:10.1111/j.1469-7610.2007.01810.x

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development, 76*, 502–517. doi:10.1111/j.1467-8624.2005.00859.x

Randell, A. C., & Peterson, C. C. (2009). Affective qualities of sibling disputes, mothers' conflict attitudes, and children's theory of mind development. *Social Development, 18*, 857–874. doi:10.1111/j.1467-9507.2008.00513.x

Razza, R. A., & Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology, 30*, 332–343. doi:10.1016/j.appdev.2008.12.020

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly, 30*, 131–147.

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology, 34*, 161–174. doi:10.1037/0012-1649.34.1.161

Ruffman, T., Perner, J., & Parkin, L. (1999). How parenting style affects false belief understanding. *Social Development, 8*, 395–411. doi:10.1111/1467-9507.00103

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*, 734–751. doi:10.1111/1467-8624.00435

Sabbagh, M., Xu, F., Carlson, S., Moses, L., & Lee, K. (2006). Development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17*, 74–81. doi:10.1111/j.1467-9280.2005.01667.x

Schneider, W., Lockl, K., & Fernandez, O. (2005). Interrelationships among theory of mind, executive control, language development, and working memory in young children: A longitudinal analysis. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 259–284). Mahwah, NJ: Erlbaum.

Shallice, T. (1982). Specific impairments of planning. *The Neuropsychology of Cognitive Function, 298*, 199–209. doi:10.1098/rstb.1982.0082

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684. doi:10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541. doi:10.1111/j.1467-8624.2004.00691.x

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). Developmental cognitive neuroscience. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1459–1474

“Two for Flinching”: Children’s and Adolescents’ Narrative Accounts of Harming Their Friends and Siblings

Holly Recchia
Concordia University

Cecilia Wainryb and Monisha Pasupathi
University of Utah

This study investigated differences in children’s and adolescents’ experiences of harming their siblings and friends. Participants ($N = 101$; 7-, 11-, and 16-year-olds) provided accounts of events when they hurt a younger sibling and a friend. Harm against friends was described as unusual, unforeseeable, and circumstantial. By contrast, harm against siblings was described as typical, ruthless, angry, and provoked, but also elicited more negative moral judgments and more feelings of remorse and regret. Whereas younger children were more self-oriented with siblings and other-oriented with friends, accounts of harm across relationships became somewhat more similar with age. Results provide insight into how these two relationships serve as distinct contexts for sociomoral development.

It has long been recognized that children’s close relationships with other children are fundamental contexts for their moral development (Piaget, 1932; Sullivan, 1953). Through their shared histories with familiar others, children learn about moral concepts such as rights, justice, benevolence, reciprocity, and trust (Bukowski & Sippola, 1996; Damon, 1977). Conflicts with peers may provide key opportunities for reflecting on these moral lessons, as children strive to find ways to balance concerns with their own desires and perspectives with their recognition of others’ divergent needs and understandings (Pasupathi & Wainryb, 2010a; Wainryb, Brehl, & Matwin, 2005). Yet we contend that children’s close relationships with different types of agemates may provide distinct opportunities for struggling with such issues. Indeed, it has been proposed that children demonstrate a particular moral concern for their friends (Youniss & Smollar, 1985). In contrast, less attention has been paid to children’s sibling relationships as contexts for moral development. Perhaps this is not surprising: Anecdotal descriptions of children’s sibling relationships as characterized by a “devastating lack of inhibition” (Dunn, 1984) or as “emotionally charged with murderous tension” (Bank & Kahn, 1982) imply that sibling interactions may not be an ideal training ground for morally upstanding behavior.

Nevertheless, we propose that everyday experiences of conflict with both friends and siblings may serve as contexts for moral development, albeit in different ways. Due to the unique and evolving provisions of children’s relationships with friends and siblings, we argue that opportunities for moral development implicated in the perpetration of harm against these two partners are likely to be different, and also to change with age. Therefore, the goal of this study was to examine the features of 7- to 16-year-olds’ narrative accounts of harming their friends and siblings, and thus to provide a window into the unique ways in which children and adolescents make sense of their distinct experiences of harm in particular close relationships.

Making Sense of Hurting or Upsetting Others

Being a moral person is typically equated with engaging in good deeds and refraining from hurting others. Yet in the course of their enduring relationships—even those with cherished playmates—children will inevitably act in ways that hurt or upset their peers, as their motivations and cognitions come

---

This research was supported by a postdoctoral fellowship from the Social Sciences and Humanities Research Council of Canada to the first author. We would like to thank the Catholic Diocese of Salt Lake City and the Murray City School District for their assistance with recruitment, Jessica Robinson for her help with coding the data, and especially the participating children and adolescents, without whom this study would not have been possible. Portions of this study were presented at the meeting of the Jean Piaget Society (2011).

Correspondence concerning this article should be addressed to Holly Recchia, Department of Education, Concordia University, 1455 de Maisonneuve West, Montreal, QC H3G 1M8, Canada. Electronic mail may be sent to hrecchia@education.concordia.ca.

© 2013 The Authors
Child Development © 2013 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0024
DOI: 10.1111/cdev.12059

into conflict with those of others. In this respect, developing a mature sense of moral agency requires acknowledging and making sense of the fact that sometimes we engage in behaviors while knowing (or suspecting) that such acts may result in harm to others. It also requires the ability to recognize that our behaviors can sometimes result in *unanticipated* harm because of unexpected clashes between our own and others' desires or interpretations. Put another way, harm can certainly result from punching and name calling, but also from more or less foreseeable goal conflicts (e.g., She wanted me to play with her but I wanted to go to the movies) as well as from unforseeable misunderstandings (e.g., I said something that made her think I don't like her, even though I *do* like her). Because children's interactions with friends and siblings tend to be quite distinct, these two relationships may provide opportunities for making sense of different experiences of harm.

Although children certainly have much to learn about the moral world from experiences in which they are the victims of others' harmful actions (e.g., Horn, 2006; MacEvoy & Asher, 2012; Wainryb et al., 2005), their attempts to make sense of experiences in which they themselves have caused harm to others may provide unique opportunities for reflecting on moral lessons. This study extends previous research that has shown that in the process of struggling to reconcile the experience of having hurt another person with their view of themselves as good people, children further their understandings of themselves and others as moral agents (Pasupathi & Wainryb, 2010a; Wainryb et al., 2005). For instance, past research reveals that when children and adolescents describe events in which they have perpetrated harm against a peer, they describe emotional and psychological consequences for the victim but simultaneously include references to their own justifiable intentions or mitigating circumstances (Wainryb et al., 2005). In line with this dual focus on self and other, children's evaluations of their own harmful actions tend to be varied; although some evaluate their behavior negatively, many others provide mixed evaluations, judging at least some aspects of their behavior to be acceptable. Overall, these data demonstrate that children's narrative accounts of their own harmful behavior may be a powerful window into how children's experiences undergird their sociomoral development.

Building on past research that broadly considered children's accounts of harm against peers in general, we propose that the manner in which children make sense of their experiences of harming agemates may be distinct, depending on the *particular* interpersonal relationship in which it occurs. In other words, we suggest that it is critical to consider how children's experiences and judgments of harm are embedded in the context of distinct relationship histories with others (Dunn, 1993; Reis, Collins, & Berscheid, 2000; Slomkowski & Killen, 1992). This proposition follows from previous research based on social domain theory (see Smetana, 2006) that underscores the ways in which children's reasoning about moral events is framed by the specific dimensions of their interpersonal relationships with others (friends and nonfriends, parents and children, men and women, in-group and out-group members; e.g., Horn, 2006; Slomkowski & Killen, 1992; Smetana & Asquith, 1994; Wainryb & Turiel, 1994). With this in mind, to inform our hypotheses, we summarize what is known about the features of children's relationships with siblings and friends, with an emphasis on the distinctive conflict processes that have been observed in each relationship. Furthermore, to address how age might moderate differences between children's narrative accounts of harming their friends and siblings, we also review how each relationship has been argued to change across development.

Relationships With Friends and Siblings: Implications for Conflict Processes

Children's relationships with close friends and siblings each have particular characteristics that differentiate them from relationships with peers in general, albeit in different ways, and that account for children's distinct conflict behaviors with these two relationship partners. Specifically, friendships are voluntary relationships defined by reciprocity and mutuality, and thus exhibit higher quality and intimacy than relationships with siblings or other same-aged peers (e.g., Buhrmester, 1992). Due to the voluntary nature of friendships, conflict may have especially serious consequences for friends as it can potentially result in the end of the relationship (Bukowski & Sippola, 1996). In line with this, friends deal with conflicts more constructively than peers who are not friends (Fonzi, Schneider, Tani, & Tomada, 1997; Vespo & Caplan, 1993), as well as siblings (DeHart, 1999; Raffaelli, 1997); friends are especially likely to use conciliatory strategies, to provide explanations, and to reach compromise resolutions on their own, without requesting the intervention of third parties.

In contrast to friendships, sibling relations during childhood are involuntary. For this reason, conflict may pose fewer risks for siblings than friends because there is no danger of ending the relationship altogether (Vandell & Bailey, 1992). Thus, although sibling relationships can certainly be intimate, playful, and loving, interactions between brothers and sisters are also known for their intensity of conflict. As compared to peers, and especially friends, conflicts between siblings are more likely to be characterized by power-assertive strategies and a lack of reasoning, to escalate to aggression, to be left unresolved, and to elicit outside intervention (DeHart, 1999; Dunn, Slomkowski, Donelan, & Herrera, 1995; Laursen & Collins, 1994; Laursen, Finkelstein, & Betts, 2001). Furthermore, although siblings are close in age, there are inherent developmental and role differences between older and younger children within a dyad. Thus, unlike friends, older and younger siblings within the dyad take on complementary roles, with older siblings exhibiting more control over interactions (Howe & Recchia, 2008). That is, older siblings act as caretakers, teachers, and protectors to their younger counterparts, but also initiate (and win) more conflicts and engage in more aggressive behavior.

Developmental Differences in Sibling Relationships and Friendships

The above body of research demonstrates that conflict processes between siblings and friends are distinct, but there is surprisingly little research that directly examines whether these relationship effects for conflict processes are of the same magnitude throughout childhood and adolescence. Although methodological discrepancies between studies at different ages make comparisons problematic (see Laursen et al., 2001), research conducted separately with young children and adolescents reveals little developmental change in how conflict processes differ between siblings and friends (cf. relationship effects for aggressive conflict strategies and outsider intervention reported by DeHart, 1999; Raffaelli, 1997).

Nevertheless, it is known that children's conceptions of sibling relationships and friendships change differentially with age. Specifically, relationships with friends become increasingly based on mutual understanding, trust, and intimate disclosure as children get older, and as they enter adolescence, young people begin to recognize that these relationships can withstand and overcome conflict (Buhrmester, 1992; Selman, 1980; Youniss, 1980). Conversely, younger children may perceive friendships as more fragile and may therefore be more likely to subordinate their own needs and desires to those of their friends (Komolova & Wainryb, 2011). In contrast, sibling relationships are described as becoming less intense and conflictual in adolescence, as teenagers place more emphasis on relationships outside of the family context (Kim, McHale, Osgood, & Crouter, 2006). At the same time, relationships between older and younger siblings become more egalitarian with age, and some studies report that sibling relationships remain a source of support and intimacy in the adolescent years (Buhrmester, 1992; Kim et al., 2006; Updegraff, McHale, & Crouter, 2002). Overall, this research seems to suggest that in some respects, perceptions of conflict with friends and sibling may converge with age, as adolescents increasingly appreciate that friendships can withstand conflict, whereas sibling relationships are simultaneously becoming more symmetrical and less contentious.

Moral Dimensions of Children's Conflicts With Friends and Siblings

Given apparent differences in the features of children's conflicts with siblings and friends, what is known about how these differences may be implicated in children's moral development? In fact, there is very little research explicitly contrasting children's moral judgments of conflict in these two relationships. The few studies examining relationship effects for children's moral evaluations of aggression reveal that harmful acts are uniformly evaluated critically, regardless of whether they are committed against siblings or peers (Astor, 1994; Shantz & Pentz, 1972). In contrast, Dunn (1993) refers to unpublished data suggesting that children evaluate transgressions against friends more negatively than transgressions against siblings. Indeed, it has been widely argued that the characteristics of children's friendships (i.e., mutual positive regard, reciprocity, and fragility) may be linked to the particular salience of moral concerns in this relationship (e.g., Bukowski & Sippola, 1996). Research contrasting children's interpretations of harm against friends and nonfriends also suggests a recognition of the unique moral status of this relationship (Slomkowski & Killen, 1992; Walton & Brewer, 2001).

Observations of the singularly aggressive and emotionally volatile nature of sibling interactions also belie the apparent similarities suggested by children's explicit evaluations of hypothetical instances of harm in these two relationships. As noted above, children's sibling interactions are known to be affectively intense and unrestrained, and can be characterized by escalating cycles of aggression (Patterson, 1986). In fact, brothers and sisters (perhaps especially *younger* brothers and sisters) appear to have special status in children's lives as causes of their anger (Hughes & Dunn, 2002; Raffaelli, 1997; Recchia & Howe, 2010). Studies examining children's attributions of blame for sibling conflicts are also informative: Children frequently perceive their brother or sister to be solely responsible for their fights (Recchia & Howe, 2010; Wilson, Smith, Ross, & Ross, 2004), although this tendency may diminish with age (e.g., Raffaelli, 1997). This latter finding provides converging evidence for our hypothesis (noted above) that descriptions of harm against siblings and friends may converge with age. However, as a whole, this research suggests that children may behave more impulsively and demonstrate more self-interest and less moral concern with their siblings than their friends.

The Current Study

This study examined children's and adolescents' narrative accounts of harm against friends and younger siblings to provide insight into their distinctive descriptions, explanations, and evaluations of harm in these two close relationships. Children's relationships with their younger (rather than older) siblings were selected as the focus of study in an effort to increase ecological validity; as noted above, research demonstrates that older siblings more often transgress against their younger siblings than vice versa (Ross, Filyer, Lollis, Perlman, & Martin, 1994).

We anticipated that children's accounts of transgressions against siblings and friends would each implicate fundamentally moral concerns such as fairness and welfare (e.g., Dunn, 1993). Nevertheless, we expected that their descriptions and evaluations of morally laden experiences would differ systematically across relationships. Specifically, to capture differences between children's understandings of transgressions against friends and siblings, we coded various aspects of how children represented the harmful acts themselves, the reasons underlying them, the context in which they occurred, and their moral implications. We also assessed dimensions that were intended to reveal children's representations of the victims' perspectives, of victims' and others' responses to harm, as well as the extent to which the narrators undermined or invalidated the victim's point of view.

Although research on this issue is limited, we were able to formulate a number of specific hypotheses based on the above literature examining children's construals of conflicts with siblings or friends (e.g., Laursen et al., 2001; Raffaelli, 1997). First, we expected that the *types* of harm against siblings would more often implicate aggression, as well as property-related issues, whereas harm against friends would be more often based on relationship-oriented concerns (e.g., betrayal, exclusion). Furthermore, with regard to the *nature* of the harm, we expected harm against siblings to be described as more provoked and less unilateral (i.e., occurring in a series of mutually harmful acts), as well as more impulsive (i.e., due to anger or disinhibition), intentional, or ruthless (i.e., demonstrating a lack of concern for the other). Along the same lines, we expected harm against friends to be described as incongruent with broader *relationship histories*, whereas harm against siblings would be more consistent with the overall tenor of the relationship. With respect to children's relative evaluations of the *validity* of their own and others' perspectives, we anticipated that narrators would selectively invalidate their siblings' points of view. Finally, we expected friends' *responses* to harm to be described as more reasoned (e.g., expressing) or avoidant (e.g., withdrawing), whereas siblings' responses to harm would be more intense or aggressive (e.g., crying, escalating), and that third parties would be more likely to get involved in conflicts between siblings than between friends.

Based on past research examining age differences in children's narrative accounts of their experiences (e.g., Pasupathi & Wainryb, 2010b; Wainryb et al., 2005), we expected that older children's accounts would be less concrete (e.g., less often based on property-related harms) and more psychological (e.g., referring more frequently to perspectives). More directly relevant to this study, we also expected narrative accounts of harm against friends and siblings to change differentially with age. Specifically, due to adolescents' increased confidence in their friendships' ability to withstand conflict (implying age-related increases in children's consideration of their *own* needs in addition to their friends' needs; Komolova & Wainryb, 2011) and the decreasing contentiousness of sibling relationships (implying age-related increases in children's consideration of their *siblings'* needs), we anticipated that narrative accounts of harm against siblings and friends would become more similar with age. More broadly, this hypothesis is also in line with theory and research suggesting developmental increases in children's ability to coordinate consideration of their own and others' needs as they reflect on social problems (Ross, Siddiqui, Ram, & Ward, 2004; Selman, 1980).

Method

Participants

Participants were recruited through advertisements in schools, day cares, community centers, and summer camps, as well as via word of mouth, in a mid-sized city in the Western United States. To be eligible for the study, children had to have at least one younger sibling, and the two children had to be born less than 4 years apart ($M_{\text{age gap}}$ = 2.46 years). The final sample included a total of 101 participants, including thirty-four 7-year-olds (M_{age} = 7.28 years, range = 6.05–8.14), thirty-three 11-year-olds (M_{age} = 11.10 years, range = 10.0–12.11), and thirty-four 16-year-olds (M_{age} = 16.10 years, range = 15.00–17.19). An additional two male participants (aged 7 and 11) were excluded because they could not remember a time that they had hurt or upset a friend. Each age group included approximately equal numbers of girls and boys (20/34 girls, 16/33 girls, and 16/34 girls, respectively). The sample was primarily Caucasian (83%), with the remaining children representing a variety of ethnic groups (African American, Hispanic, Asian, Middle Eastern, Native American, and mixed descent). Parents provided written informed consent; children provided written assent to all procedures. Each child received a movie gift certificate in appreciation for his or her participation.

Procedure

Data for this study were drawn from a larger investigation of children's moral development; only procedures relevant to the current investigation will be described here.

In individual interviews (conducted either at a university lab or the child's home), children were privately asked to provide narrative accounts of (a) a time when they hurt or upset a younger sibling, and (b) a time when they hurt or upset a friend ("Tell me about a time when you did or said something that ended up hurting or upsetting *your brother/sister* OR *one of your good friends*. Try to tell me everything you remember about that time."). The order of these two narrative elicitations was counterbalanced within age and gender. Participants were asked to choose events that were important to them and that they remembered well. If children nominated a generic or recurrent event (e.g., "I always take his stuff"), they were asked to provide an account of one specific episode. The interviewer encouraged participants to continue speaking by using general prompts ("uh huh…", "and …?"), or by repeating verbatim what the child said ("so then you left without her…"). When the child appeared to have come to the end of her or his narrative, the interviewer asked, "Is there anything else you remember about that time?" This procedure ensures that the interviewer provides no cues for the types of content (references to reasons, emotions, etc.) that participants include in their narrative accounts (Wainryb et al., 2005). Interviews were audiotaped and transcribed verbatim for analysis.

Coding

Coding categories were adapted from research on children's narrative accounts of their own transgressive behavior (Wainryb et al., 2005); revisions and additions to existing coding schemes were informed by research examining children's distinct construals of relationships with siblings and friends (e.g., Raffaelli, 1997; Recchia & Howe, 2010). Coders (i.e., the first author and a second naïve coder) first discussed the categories and their definitions and then trained to criterion by jointly coding a subset of 10% of the narratives; interrater reliability was then established on an additional 20% of the narratives. Disagreements were resolved via discussion and consensus. Cohen's kappas are reported below.

Types of harm. Each narrative was coded for presence of six types of harm (κ = .92): (a) harm resulting from *offensive behavior* (e.g., yelling, insulting, teasing), (b) *physical* harm (e.g., hitting, kicking, pushing), (c) *property*-related harm (e.g., failing to share, property destruction), (d) *relationship*-based harm (e.g., physical or psychological separation, trust violation), (e) harm resulting from the victim's *blocked goal* (e.g., incompatible plans for play, winning a game), or (f) harm resulting from *honesty/insensitivity* (e.g., saying "Do I look fat in this?" to an overweight friend, or that a sibling's writing assignment was only "pretty good").

Perpetrator's reasons for harm. Each narrative was coded for the presence of eight types of reasons for the harmful behavior (κ = .93): (a) *unintentional* (i.e., accidents, jokes, or negligence), (b) *benevolent* (i.e., prosocial intent), (c) *emotional/impulsive* (i.e., driven by anger or a lack of control), (d) pursuit of an instrumental *goal* (e.g., to watch television, to pass a test), (e) *malicious* (i.e., retributive, intent to harm), (f) *perpetrator's misunderstanding* (i.e., mistaken assumptions, lack of knowledge), (g) *provocation* (i.e., a response to an offensive or irritating behavior by the victim), (h) *extenuating circumstances* (i.e., behavior resulting from factors beyond the perpetrator's control, such as parents' directives).

Ruthlessness. In addition to explicit references to the perpetrator's malicious intent, we also coded more implicit descriptions of ruthless behavior resulting in harm ($\kappa = .82$). Specifically, for each narrative, we noted whether the narrator's behavior demonstrated an apparent disregard for the victim by deliberately and reflectively engaging in an action that was clearly understood to be intrinsically harmful (e.g., "I was hitting her Barbies across the room with a golf club"), or using harmful means in service of a goal that could be accomplished in a much more straightforward or benign way (e.g., "He was chewing with his mouth open, so I called him a fag").

Mutuality. For each narrative, we coded whether the narrator described the event as one of unilateral harm (i.e., he or she was the sole perpetrator) or whether both protagonists engaged in mutually harmful behavior ($\kappa = .79$).

Relationship context. For each narrative, we coded whether the narrator referred to the harmful act as being *congruent* with the overall history of the relationship with the victim (e.g., "This happens all the time"), *incongruent* with the relationship (e.g., "Normally I don't do that"), or whether they made *no reference* to the relationship context of the harm ($\kappa = .81$).

Moral concerns surrounding harm. Each narrative was coded for spontaneous references to moral evaluations or moral emotions surrounding the harmful behavior ($\kappa = .79$). Specifically, this category encompassed negative evaluations of the act (e.g., "It was really mean"), references to the absurdity or senselessness of the act or the conflict (e.g., "We fight about stupid things that don't even matter"), and references to remorse or regret (e.g., "I felt crummy about it").

References to the victim's perspective. For each narrative, we coded whether the narrator made reference to the victim's perspective explaining why he or she was hurt by the narrator's behavior ($\kappa > .76$). Specifically, we coded any reference to the victim's conflicting *motivations* (e.g., "He wanted to play with us"), *cognitions* (e.g., "He thinks it's not fair"), and *emotions* (e.g., "He was feeling kind of cranky").

Victim's misunderstanding of the perpetrator's behavior. In addition to coding overall references to the victim's perspective, we also coded instances when narrators made reference to the *victim's misunderstanding of the perpetrator's behavior* that increased the potential for harm (e.g., "I was being sarcastic, and she took it the wrong way"; $\kappa = .93$).

Invalidating the victim's perspective. For each narrative, we coded whether the narrator implicitly or explicitly *invalidated* the victim's point of view ($\kappa = .86$). Specifically, this encompassed instances when the narrator described the victim as unreasonable or incomprehensible (e.g., "He makes no sense"), suggested that the victim's reaction to the narrator's behavior was overstated (e.g., "She overreacted and threw a fit"), or expressed a sense of righteous indignation (e.g., "I'm like—do you want my help or not?").

Emotional consequences for the victim. Each narrative was coded for whether the narrator referred to an emotional consequence of the harm for the victim; these consequences were further coded specifically for references to victims' *anger* (e.g., mad, pissed off, frustrated) or *sadness* (e.g., hurt feelings, felt bad; $\kappa = .88$).

Responses to harm. Each narrative was coded for the presence of six possible responses to harm, including those of the victim as well as third parties ($\kappa = .95$): (a) *victim avoiding* (e.g., running away), (b) *victim escalating* (e.g., hitting, yelling), (c) *victim expressing* a conflicting position (e.g., verbally expressing a blocked goal), (d) *victim crying* (e.g., "She started tearing up"), (e) *victim reacting with anger* when such responses did not include escalation (e.g., "She glared at me"), and (f) *adult intervention* (e.g., "My mom said stop").

Results

Analyses were conducted separately for each type of narrative content as a function of relationship (friend, sibling), age (7-, 11-, and 16-year-olds), and gender, with relationship as a repeated measure. Analysis of variance (ANOVA)-based procedures were used because they have been shown to be more appropriate for analyzing this type of data than are log linear-based procedures, as the latter run into a distinct estimation problem (see Wainryb, Shaw, Laupa, & Smith, 2001). For each significant omnibus effect, effect size is reported as partial eta-squared. Bonferroni corrections (with an alpha level of $p < .05$) were used for all post hoc pairwise comparisons.

How Do Narrators Describe Their Own Harmful Behaviors With Friends and Younger Siblings?

Types of harm. Narratives could include multiple categories of harm because some participants reported engaging in multiple harmful behaviors (19% of narratives included references to two or more harmful acts). As such, these codes were analyzed using a Relationship × Age × Gender multivariate analysis of variance (MANOVA) with the six types of harmful acts as dependent variables. This analysis revealed multivariate effects for relationship, Wilk's $\lambda = .67$, $\eta^2 = .33$, age, $\lambda = .77$, $\eta^2 = .12$, and the Relationship × Age × Gender three-way interaction, $\lambda = .79$, $\eta^2 = .11$.

Follow-up ANOVAs revealed effects of relationship on offensive behavior, $F(1, 95) = 15.85$, $\eta^2 = .14$; property harm, $F(1, 95) = 13.64$, $\eta^2 = .13$; relationship harm, $F(1, 95) = 13.24$, $\eta^2 = .12$; and honesty/insensitivity, $F(1, 95) = 9.81$, $\eta^2 = .09$. Results are reported in Table 1. Narratives about harm against siblings were more likely to include references to offensive behavior and property-related harms, whereas narratives about harm against friends were more likely to include references to relationship-based harms and honesty/insensitivity.

In addition, follow-up ANOVAs revealed age effects for offensive behavior, $F(2, 95) = 3.25$, $\eta^2 = .06$; property-related harms, $F(2, 95) = 3.27$, $\eta^2 = .06$; blocked goals, $F(2, 95) = 3.64$, $\eta^2 = .07$; and honesty/insensitivity, $F(2, 95) = 3.36$, $\eta^2 = .07$. Follow-up t tests revealed that offensive behavior was significantly more likely among 16-year-olds ($M = .41$) than 7-year-olds ($M = .21$), with 11-year-olds falling in between ($M = .31$). In contrast, property-related harms occurred more frequently among 7-year-olds ($M = .34$) than 11-year-olds ($M = .15$), with 16-year-olds not significantly different from either group ($M = .21$). Harm resulting from blocked goals was more frequent among 11-year-olds ($M = .26$) than 16-year-olds ($M = .09$), with 7-year-olds not significantly different from either group ($M = .13$). Finally, harm resulting from honesty/insensitivity was more likely among 16-year-olds ($M = .18$) than 7-year-olds ($M = .04$), with 11-year-olds not significantly different from either group ($M = .08$).

Main effects of relationship and age on property-related harms were qualified by the Relationship × Age × Gender interaction, $F(2, 95) = 3.33$, $\eta^2 = .07$. A follow-up analysis of the pattern of means revealed that the relationship effect for this type of harm was only statistically significant for 7-year-old girls and 11-year-old boys. For both genders, property-related harms among 16-year-olds were equally likely to occur with siblings and friends.

Perpetrator's reasons for harm. Similar to harmful acts, participants could refer to multiple intentions underlying their harmful behavior in the same narrative. Therefore, we conducted a Relationship × Age × Gender MANOVA with the eight types of reasons as dependent variables. This analysis revealed only a significant multivariate effect for relationship, $\lambda = .73$, $\eta^2 = .27$. Specifically, follow-up ANOVAs revealed relationship effects for benevolent reasons, $F(1, 95) = 4.16$, $\eta^2 = .04$; emotional/impulsive reasons, $F(1, 95) = 12.67$, $\eta^2 = .12$; provocation, $F(1, 95) = 4.45$, $\eta^2 = .05$; and extenuating circumstances, $F(1, 95) = 15.46$, $\eta^2 = .14$. Consistent with hypotheses, whereas benevolent reasons and extenuating circumstances were more frequent in accounts of harm against friends, emotional/impulsive reasons and provocation were described more often with siblings (see Table 1).

Ruthlessness. A Relationship × Age × Gender ANOVA with ruthlessness as the dependent variable revealed an effect of relationship, $F(1, 95) = 11.71$, $\eta^2 = .11$. As expected, narrators described behaving more ruthlessly with siblings than with friends (see Table 1).

Mutuality of harm. A Relationship × Age × Gender ANOVA with mutuality as the dependent variable revealed a Relationship × Age interaction, $F(2, 95) = 3.43$, $\eta^2 = .07$. Consistent with age-related predictions, post hoc t tests revealed that 7-year-olds described mutual harms more often with siblings than with friends, whereas relationship effects were not significant for 11-year-olds or 16-year-olds (see Table 3).

Relationship context of harm. A Relationship × Age × Gender MANOVA with congruent and incongruent relationship contexts as dependent variables revealed a multivariate effect of relationship, $\lambda = .73$, $\eta^2 = .27$. Follow-up ANOVAs revealed that relationship effects were significant for both congruent, $F(1, 95) = 29.55$, $\eta^2 = .24$, and incongruent relationship contexts, $F(1, 95) = 13.20$, $\eta^2 = .12$. Whereas sibling harm was more often congruent with the relationship, harm against friends was more often incongruent with the relationship (see Table 1).

Moral concerns surrounding harm. A Relationship × Age × Gender ANOVA with moral concerns as the dependent variable revealed an effect of relationship, $F(1, 95) = 4.75$, $\eta^2 = .05$. References to moral concerns were more frequent with siblings than with friends (see Table 1).

Descriptions of harmful behavior: Summary of key findings. Consistent with hypotheses, harm against friends was based on relationship-oriented concerns and relatively benign behaviors such as honesty/insensitivity, and harm against siblings was due to more objectively offensive behavior or property-related issues. In turn, in their accounts of harming their friends, children described more benevolent goals and extenuating circumstances, whereas harm against siblings was described as more emotional/

Table 1
Relationship Differences Between Descriptions of Harmful Behavior with Friends and Siblings

	Harm against friends *M* proportion of narratives (*SE*)	Harm against siblings *M* proportion of narratives (*SE*)	Overall *M* proportion of narratives (*SE*)
Types of harm			
Offensive behavior	.19 (.04)$_a$	.43 (.05)$_b$	.31 (.03)
Physical	.15 (.03)	.18 (.04)	.16 (.03)
Property	.13 (.03)$_a$	.33 (.05)$_b$	.23 (.03)
Relationship	.29 (.05)$_a$	.09 (.03)$_b$	.19 (.03)
Blocked goal	.19 (.04)	.12 (.03)	.16 (.03)
Honesty/insensitivity	.16 (.04)$_a$	.04 (.02)$_b$	.10 (.02)
Perpetrator's reasons			
Unintentional	.12 (.03)	.16 (.04)	.14 (.02)
Benevolent	.08 (.03)$_a$	.02 (.01)$_b$	.05 (.02)
Emotional/impulsive	.11 (.03)$_a$	.28 (.05)$_b$	.20 (.03)
Goal	.32 (.05)	.36 (.05)	.34 (.03)
Malicious	.02 (.01)	.05 (.02)	.04 (.01)
Perpetrator's misunderstanding	.12 (.03)	.09 (.03)	.10 (.02)
Provocation	.34 (.05)$_a$	.47 (.05)$_b$	.40 (.04)
Extenuating circumstances	.16 (.04)$_a$	.01 (.01)$_b$	.09 (.02)
Ruthlessness	.12 (.03)$_a$	.30 (.05)$_b$	.21 (.03)
Mutuality	.46 (.05)	.51 (.05)	.49 (.04)
Relationship context			
Congruent	.20 (.04)$_a$	.54 (.05)$_b$	.37 (.03)
Incongruent	.17 (.03)$_a$	.03 (.02)$_b$	.10 (.02)
Moral concerns	.15 (.04)$_a$	.26 (.04)$_b$	.21 (.05)

Note. Means in the same row are labeled with different subscripts when post hoc Bonferroni tests revealed significant differences at $p < .05$. Proportions within a column may sum to greater than 1.0 for types of harm and reasons for harm because it was possible for multiple categories to be coded for the same narrative.

impulsive, provoked, ruthless, and typical of the relationship. Interestingly, children were more likely to spontaneously refer to moral concerns in their accounts of harming their siblings. Finally, consistent with hypotheses, relationship differences in children's descriptions of the mutuality of harm were especially pronounced among 7-year-olds.

How Do Narrators Describe the Victim's Perspective and Responses to Harm?

References to the victim's perspective. Participants could refer to multiple aspects of victims' overall perspectives in the same narrative. As such, we conducted a Relationship × Age × Gender MANOVA with references to the victim's conflicting goals, cognitions, and emotions as dependent variables. This analysis revealed multivariate effects of age, $\lambda = .86$, $\eta^2 = .07$, and gender, $\lambda = .92$, $\eta^2 = .08$. Follow-up ANOVAs revealed an age effect on references to the victim's cognitions, $F(2,\ 95) = 3.42$, $\eta^2 = .07$. Specifically, references to the victim's cognitions were more frequent among 16-year-olds ($M = .41$) than 7- or 11-year-olds (both $Ms = .23$). Similarly, ANOVAs revealed a gender effect for victim's cognitions, $F(1,\ 95) = 4.34$, $\eta^2 = .04$. Girls ($M = .35$) referred to the victims' cognitions more often than boys ($M = .22$).

Victim's misunderstanding of the perpetrator's behavior. A Relationship × Age × Gender ANOVA examining narrators' references to the victim's misunderstanding of the perpetrator's behavior revealed only a main effect of relationship, $F(1,\ 95) = 6.34$, $\eta^2 = .06$. These references were more frequent for friends than for siblings (see Table 2).

Invalidating the victim's perspective. A Relationship × Age × Gender ANOVA with statements *invalidating the victim's perspective* as the dependent variable revealed a Relationship × Age interaction, $F(2,\ 95) = 3.62$, $\eta^2 = .07$. In line with hypotheses, follow-up t tests revealed that 7-year-olds were more likely to invalidate their younger sibling's perspective than that of their friend. In contrast, 11-year-olds and 16-year-olds were equally likely to

Table 2
Relationship Differences Between Descriptions of Friends' and Siblings' Perspectives and Responses to Harm

	Harm against friends *M* proportion of narratives (*SE*)	Harm against siblings *M* proportion of narratives (*SE*)	Overall *M* proportion of narratives (*SE*)
Victim's perspective			
Goals	.47 (.05)	.49 (.05)	.48 (.03)
Cognitions	.34 (.05)	.24 (.04)	.29 (.03)
Emotions	.09 (.03)	.06 (.02)	.07 (.02)
Victim's misunderstanding of perpetrator's behavior	.18 (.04)$_a$	.06 (.02)$_b$	.12 (.02)
Invalidating the victim's perspective	.18 (.04)	.27 (.05)	.22 (.03)
Emotional consequences for the victim			
Anger	.42 (.05)	.42 (.05)	.42 (.04)
Sadness	.30 (.05)	.25 (.04)	.27 (.04)
Responses to the harm			
Victim avoiding	.28 (.04)	.18 (.04)	.23 (.03)
Victim escalating	.24 (.04)$_a$	.35 (.05)$_b$	.30 (.04)
Victim expressing	.37 (.05)$_a$	.20 (.04)$_b$	.29 (.03)
Victim crying	.13 (.03)$_a$	.25 (.04)$_b$	.19 (.03)
Victim reacting with anger	.01 (.01)$_a$	.07 (.03)$_b$	.04 (.01)
Adult intervention	.14 (.03)$_a$	.42 (.05)$_b$	.28 (.03)

Note. Means in the same row are labeled with different subscripts when post hoc Bonferroni tests revealed significant differences at $p < .05$. Proportions within a column may sum to greater than 1.0 for victims' perspectives, emotional consequences, and responses because it was possible for multiple categories to be coded for the same narrative.

invalidate their younger siblings' and friends' perspectives (see Table 3).

Emotional consequences for the victim. Narrators could refer to multiple emotional consequences. As such, we conducted a Relationship × Age × Gender MANOVA with the victim's anger and sadness as dependent variables, revealing a multivariate effect of gender, $\lambda = .91$, $\eta^2 = .09$, and a multivariate Relationship × Age interaction, $\lambda = .84$, $\eta^2 = .09$.

Follow-up ANOVAs revealed a gender effect for references to victims' anger, $F(1, 95) = 5.94$, $\eta^2 = .06$. Girls ($M = .52$) were more likely than boys ($M = .32$) to refer to anger. In turn, ANOVAs revealed a Relationship × Age interaction for references to victims' sadness, $F(2, 95) = 9.10$, $\eta^2 = .16$. *T* tests revealed that 7-year-olds described their friends as experiencing sadness more than their younger siblings, whereas 16-year-olds described their younger siblings as experiencing sadness more than their friends. In contrast, 11-year-olds were equally likely to describe their friends and siblings as sad (see Table 3).

Responses to harm. Participants could refer to multiple responses to harm. As such, we conducted a Relationship × Age × Gender MANOVA with the six types of responses to harm as dependent variables. This analysis revealed multivariate effects for relationship, $\lambda = .64$, $\eta^2 = .36$; age, $\lambda = .72$, $\eta^2 = .15$; and a Relationship × Gender interaction, $\lambda = .86$, $\eta^2 = .14$.

Follow-up ANOVAs revealed relationship effects for victim escalating, $F(1, 95) = 3.84$, $\eta^2 = .04$; victim expressing, $F(1, 95) = 7.64$, $\eta^2 = .07$; victim crying, $F(1, 95) = 4.07$, $\eta^2 = .04$; victim reacting with anger, $F(1, 95) = 4.72$, $\eta^2 = .05$; and adult intervention, $F(1, 95) = 25.18$, $\eta^2 = .21$. As expected, whereas friends were more likely to express their perspectives, siblings were more likely to escalate, cry, and react with anger, and adults were more likely to intervene with siblings than with friends (see Table 2).

In turn, ANOVAs revealed age effects for expressing perspectives, $F(2, 95) = 9.16$, $\eta^2 = .16$, and adult intervention, $F(2, 95) = 4.09$, $\eta^2 = .08$. *T* tests revealed that 16-year-olds ($M = .46$) were more likely than 11-year-olds ($M = .17$) and 7-year-olds ($M = .23$) to describe victims as responding by expressing their perspective. In contrast, 7-year-olds ($M = .40$) were more likely to refer to adult intervention than 16-year-olds ($M = .18$), with 11-year-olds not significantly different from either group ($M = .26$).

Finally, ANOVAs revealed a Relationship × Gender interaction for avoiding, $F(1, 95) = 11.90$, $\eta^2 = .11$. Whereas girls described their friends ($M = .45$) as engaging in avoidant reactions more often than their siblings ($M = .17$), boys were

Table 3

Age-Related Changes in the Differences Between Children's Accounts of Harming Their Friends and Siblings

	7-year-olds		11-year-olds		16-year-olds	
	Harm against friends *M* proportion of narratives (*SE*)	Harm against siblings *M* proportion of narratives (*SE*)	Harm against friends *M* proportion of narratives (*SE*)	Harm against siblings *M* proportion of narratives (*SE*)	Harm against friends *M* proportion of narratives (*SE*)	Harm against siblings *M* proportion of narratives (*SE*)
Mutuality of harm	.34 (.09)$_a$	.60 (.09)$_b$	.64 (.09)	.52 (.09)	.41 (.08)	.40 (.09)
Invalidating the victim's perspective	.08 (.07)$_a$	.34 (.08)$_b$	.15 (.07)	.27 (.08)	.30 (.07)	.21 (.08)
Emotional consequences for the victim (sadness)	.35 (.08)$_a$	.09 (.07)$_b$	.33 (.08)	.21 (.08)	.21 (.08)$_a$	.45 (.07)$_b$

Note. Means in the same row are labeled with different subscripts when post hoc Bonferroni tests revealed significant simple effects of relationship type at $p < .05$.

equally likely to describe their friends and siblings as avoiding (*M*s = .11 and .20, respectively).

The victim's perspective and responses to harm: Summary of key findings. Results revealed that friends, more than siblings, were described as experiencing harm due to their misinterpretations of the perpetrator's behavior. As expected, friends were described as responding to harm in more reasoned ways, whereas siblings were seen as more likely to escalate, cry, and react with anger, and adults were depicted as being more likely to intervene in sibling conflict. Also in line with hypotheses, 7-year-olds were especially likely to invalidate their sibling's perspective (relative to their friend's), and with increasing age, children became more likely to acknowledge siblings' hurt feelings.

Discussion

The goal of this study was to examine children's and adolescents' narrative accounts of harming their friends and siblings. In some ways, our results revealed considerable consistency across relationships in children's experiences of hurting or upsetting others, thus replicating overall patterns previously observed in children's narratives of perpetrating harm against agemates (Wainryb et al., 2005). For example, regardless of the relationship context in which it occurred, children rarely described hurting others for explicitly malicious reasons and tended to maintain a dual focus on their own justifiable intentions and the victim's hurt feelings. Inasmuch as these patterns are distinct from those observed in children's accounts of being victimized by others (see Wainryb et al., 2005), our results add to a body of work suggesting that children's experiences of hurting or upsetting others provide unique opportunities for moral reflection (Pasupathi & Wainryb, 2010a).

Importantly, however, we also extended this research by revealing differences in how children make sense of their own harmful behavior in two distinctive close relationships with agemates. Furthermore, this study is the first to contrast children's construals of harm in these two relationships across a wide age range. These findings build on research demonstrating that children's constructions of meanings and judgments about moral events vary as a function of their relationship histories with others (e.g., Slomkowski & Killen, 1992; Wainryb & Turiel, 1994). More specifically, analyses of participants' narratives revealed a number of descriptive and evaluative differences between children's accounts of harm in these two close relationships, as well as some intriguing age-related changes in the magnitude of these differences. Although our results revealed a few effects of gender, overall, the observed patterns were similar for boys and girls. Taken as a whole, our results suggest that children's reflections on their perpetration of harm in these two relational contexts may provide distinct but complementary opportunities for sociomoral development.

How Do Children Make Sense of Their Perpetration of Harm Against Friends?

Inasmuch as children are both invested in their friendships and motivated to protect and preserve them, children may avoid harming their friends and thus risk damaging the relationship, especially prior to adolescence. Partly for these reasons, it has been proposed that friendships may be a context in which children demonstrate a heightened moral concern for the needs of others (Bukowski & Sippola, 1996). Nevertheless, in the course of repeated interactions, children will occasionally engage in behaviors that hurt or upset their friends. As such, given the apparent inevitability of these experiences, how might children make sense of these events in ways that contribute to their moral development?

Consistent with their views of friendships as supportive and affectively positive contexts (Buhrmester, 1992), our participants often noted that incidents of harm were discrepant with the overall tenor of their friendship histories. Furthermore, the *ways* in which children described harming their friends reflected the unique provisions of this relationship. First, and as expected, our results revealed that, compared to siblings, harms against friends were more often described as based on relationship-centered concerns, such as trust and the desire for connectedness, reflecting children's strong investment in this relationship. Indeed, among our participants, the most frequent source of harm against friends was simply failing to share time or space with the other, rather than more unambiguous moral transgressions such as name calling or refusing to share material possessions. In addition, as compared to siblings, harm against friends was described as less forseeable and more often resulting from external constraints, in that it resulted from friends' interpretations of children's ambiguous, insensitive, or even benevolent (but perhaps misguided) behavior, or circumstances beyond the narrator's control. Consider the following example, which is fairly typical of children's accounts of harming friends (edited for length where indicated; all names are pseudonyms):

> I'm an athletic girl. I like go and play kick ball and one of my really good friends named Jenna, she plays with these girls who . . . are always doing a lot of athletic things . . . so I was just playing with Jenna that day and Kelsey got mad at me and thought I hated her . . . cuz I wasn't playing with her. And so then we were talking on Facebook and I told her why and we, we hanged out after. (11-year-old girl)

In this narrative, the harm both fundamentally implicates the relationship itself and is described as an unfortunate, unanticipated consequence of goal-directed behavior that is interpreted in a hurtful way by the other child. Extending previous research and theory, this pattern demonstrates how children's particular moral sensitivity to their friends may be manifested in the context of their experiences of harming others. Overall, these results suggest that even in a relationship in which they are highly attuned to the needs of others (e.g., Komolova & Wainryb, 2011), children recognize through these experiences that they cannot always anticipate how their friends will react to their actions. Therefore, we argue that harming one's friends provides children with important insights about their imperfect grasp of others' perspectives, as well as others' imperfect grasp of their own.

How Do Children Make Sense of Their Perpetration of Harm Against Siblings?

In contrast to children's friendships, the characteristics of sibling interactions are presumed to be less conducive to moral lessons. Indeed, many of the narrative accounts in our data reflect the uninhibited (and occasionally ruthless) nature of children's harmful behavior against their siblings (edited for length where indicated):

> Ben was making faces at me, and I'm not really good at faces, I'm actually really good with my hands . . . and I hit him. (7-year-old boy)
>
> I called him a stupid, mean, nasty little elf-brother! He IS pretty short. (11-year-old girl)
>
> I learned this thing from my friends, like when you make somebody flinch, you punch them twice and say "two for flinching." So I did that to her and I just kept on doing it and doing it and doing it. (11-year-old boy)
>
> I put a sign on his back that said "poke me" . . . I was like "oh, it would be really funny to put this sign on Kevin's back." (16-year-old girl)

Nevertheless, despite the apparently callous nature of some of these acts, we propose that children's experiences of engaging in behavior while knowing or suspecting that such behavior might result in harm to others may be a critical context for moral development (Pasupathi & Wainryb, 2010a, Wainryb et al., 2005). With this in mind, what are the features of children's narrative accounts of harming their siblings that might shed light on the particular insights that children could be gaining from these experiences? First, and as expected, more so than with friends, harm to siblings typically resulted from explicitly offensive behavior (e.g., teasing) or property-related violations (e.g., sharing), although it should be noted that this latter effect was qualified by age and gender. As illustrated in the examples above, harm to siblings was also more often ruthless in tone. Furthermore, also as expected, children described their harmful actions against siblings as driven by anger or a lack of control, often in response to the other's provocation. Indeed, the disinhibited nature of sibling interaction is also exemplified by siblings' responses to harm. More than friends, siblings who were the victims of harm were described as responding in emotionally expressive or explosive (rather than reasoned) ways by escalating, crying, or reacting with anger. Taken together, this pattern suggests that, even in the early school-aged years, children demonstrate an understanding of cycles of coercion: In many cases, they described how provocation by their sibling and their resultant anger led them to lash out in hurtful ways, which prompted an increasingly aggressive and emotional reaction from their sibling. In other words, children's accounts suggest that they recognize the coercive patterns that have been observed to characterize some sibling interactions (Patterson, 1986).

Even more importantly, our results suggest that not only are children capable of describing the more negative features of sibling interactions but also that they themselves judge such interaction patterns to be problematic. Specifically, alongside their acknowledgment of the occasionally ruthless nature of their behavior with their sibling, as compared to their narrative accounts of harm against friends, children were more likely to make references to moral concerns surrounding these experiences, including references to the absurdity or senselessness of conflict, negative evaluations of their own hurtful actions, and feelings of remorse or regret. To illustrate, consider the following example (edited for length, where indicated):

> I'm kind of mean to her just 'cause she's my sister, you know? Cuz I'm kind of sarcastic but I'm kind of not, and so she gets upset, and then it is sad. . . . Sometimes I tell her like, "oh just because you don't have any friends, that doesn't mean you can come and hang out with my friends" because that's something that usually gets to her, so you tend to lean towards that, to make her go away. . . . One time . . . she kept on trying to play with me and my friend, and we were yelling at her . . . We hurt her a little bit and she started crying. And so that's when we realized like, ok, you can really hurt someone, like unintentionally, so it's important to watch what you're about to say and stuff. . . . I wish it didn't happen. . . . We were really mad at her, and I don't know why, and she just really wanted someone to play with. (16-year-old girl)

In this example, the narrator depicts the transgression against her sibling as emotionally driven (i.e., based on anger) and ruthless (i.e., motivated to upset her sister so that she would go away)—both of these elements are consistent with the reckless nature of sibling harm. However, this participant also conveys a sense of clear remorse and a negative judgment of her own behavior. On its surface, this latter finding might appear to be counterintuitive: Why might these references to moral judgments and emotions be more common in narrative accounts of harm against siblings than about friends? We suggest at least two possible explanations of this pattern. First, our findings imply that children's harmful actions against their sibling are experienced in more agentic ways than with their friends; whereas harm against friends was characterized as ambiguous, unanticipated, and due to extenuating circumstances, harm against siblings was more often described as internally motivated (e.g., by anger). Thus, to the extent that children construct an understanding of their transgressions against their sibling as explicit, ruthless, and internally driven behaviors, these experiences may be more likely to pose a challenge to their views of themselves as moral people. As such, harm against one's sibling may be more accompanied by a sense of remorse or regret, and thus encourage young people to consider their actions in light of their moral values. Second, we suggest that the feedback that children receive from their siblings and parents might be particularly helpful in this regard. More specifically, awareness of the negative emotional consequences of behavior is a critical determinant of moral judgments (Arsenio, Gold, & Adams, 2006); in contrast to friends, our results demonstrate that siblings are more likely to provide emotionally intense (and apparently memorable) negative responses to children's hurtful actions. Furthermore, because parents can scaffold children's understandings of their conflicts with others (Recchia, Wainryb, & Pasupathi, 2011; Thompson, 2006), to the extent that adults are described as intervening more frequently in situations involving siblings than those involving friends, they may be helping children to draw meanings from experiences with their brothers and sisters. At any rate, these findings underscore that children's experiences of conflict with their sibling should not be overlooked as relational contexts for moral development, in that they appear to provide unique opportunities for moral learning in their own right.

How Do Children's Understandings of Harm Against Friends and Siblings Change With Age?

Our results revealed a number of age-related changes in children's narrative accounts of harm that confirm well-established developmental trends in children's narrative accounts of their experiences (e.g., Pasupathi & Wainryb, 2010b). Specifically, property-related harms were particularly frequent among 7-year-olds (e.g., sharing toys), and harms resulting from blocked goals (e.g., conflicting plans for play) occurred especially among 11-year-olds. In contrast, harms resulting from offensive behavior and honesty/insensitivity were described more frequently with age. Consistent with this pattern, with age, narrators were also increasingly likely to refer to the cognitions of the victim. Taken together, these results suggest that experiences of harm became more psychologically based and less concrete with increasing age.

More germane to the novel questions addressed in this study, we also examined how the magnitude of relationship differences changed with age. Indeed, it is known that the provisions of sibling relationships and friendships each change with development; with age, children becoming increasingly confident that their friendships can withstand and overcome disagreement (Komolova & Wainryb, 2011), whereas their sibling relationships are simultaneously becoming less contentious (e.g., Buhrmester, 1992). Furthermore, with age, children become increasingly capable of coordinating multiple perspectives in ways that permit the simultaneous consideration of their own and others' positions in the context of social conflict (Ross et al., 2004; Selman, 1980). Thus, although little research has examined developmental differences in children's construals of conflicts with siblings and friends, we expected that children's accounts of harming their siblings and friends would tend to converge with age.

Overall, it is important to underscore that our results revealed considerable continuity across development with regard to the distinctiveness of children's accounts of harming their siblings and friends. Results revealed many relationship effects that were consistent across age, implying that differences between relationships with siblings and friends may be as salient for adolescents as they are for younger children. Possibly, research documenting these patterns across the transition into early adulthood may reveal more robust declines in the emotional intensity of conflict between siblings than those observed here (e.g., Conger & Little, 2010).

Nevertheless, in the three instances when relationship effects were moderated by age, our results were consistent with hypotheses. Specifically, in contrast to 11- and 16-year-olds, 7-year-olds more frequently described their harmful acts against siblings as occurring in the context of a mutual series of oppositional behaviors than their harmful acts against friends. Along the same lines, 7-year-olds (but not older participants) selectively invalidated the conflict perspective of their sibling by dismissing it as unreasonable or incomprehensible, whereas they were less likely to do so with friends. Finally, as a consequence of their own harmful behavior, 7-year-olds were more likely to refer to their friends' sadness than their siblings' sadness. Interestingly, the opposite pattern was observed among 16-year-old participants, who were more likely to refer to their siblings' sadness than that of their friends.

These findings demonstrate that by the early school-aged years, children have differentiated views of their close relationships that provide distinct opportunities for learning about negotiating interactions with others. More specifically, based on these results, we suggest that young children's conflicts with friends may be particularly conducive to considering the needs of others: that one's own actions can make others sad and that others' experience of sadness is legitimate and comprehensible. In contrast, the provisions of young children's sibling relationships appear to be linked to quite different construals: that sometimes one's own perspective has unique validity and that one's own actions need not be considered in unilateral terms but rather that the roles of culprits and victims are not always differentiable. In this respect, experiences of harm against siblings and friends may provide complementary opportunities for young children to consider the needs of self and other.

Interestingly, our results also suggest that with increasing age, accounts of harming siblings and friends become somewhat more similar. More specifically, given adolescents' changing conceptions of friendships, it may become less threatening to acknowledge that their friends' perspectives are sometimes invalid and in these cases, that it is justified to stand one's ground. Similarly, although adolescents continue to behave in disinhibited ways with their siblings, they also increasingly recognize the hurtful consequences of this behavior. When looked at as a whole, our results provide intriguing and novel insight into children's evolving experiences of harm in their varied close relationships.

Conclusions

Our study was designed to investigate children's and adolescents' construals of harm against siblings and friends across a wide age range. With this goal in mind, we asked children to furnish narrative accounts that captured their spontaneous descriptions and evaluations of their own past experiences. Although our interest was not in determining the objective or observable features of conflicts with siblings and friends, it is important to note that our methods do not reveal the actual range of conflict experiences that children might have with their siblings and friends. Indeed, as with all retrospective reports, our findings may be partially driven by selection effects; for this reason, children's reasoning and judgments about hypothetical conflict scenarios involving friends and siblings may provide a useful complement to this work. In future research, it would also be interesting to examine within-person heterogeneity in accounts of sibling and friendship conflict; studies eliciting accounts of a variety of different conflicts with siblings and friends (e.g., welfare vs. justice; victim vs. perpetrator; resolved vs. unresolved) may provide further insight into children's and adolescents' experiences in different relationships. Moreover, it would be useful to determine whether the pattern of results for siblings would generalize to descriptions of harm perpetrated against older (rather than younger) brothers or sisters, given well-documented differences in older and younger siblings' roles in the family (Howe & Recchia, 2008); this is a critical issue for future work, given that status differences are known to influence children's moral reasoning and judgment (e.g., Horn, 2006; Wainryb & Turiel, 1994).

Nevertheless, our study makes a number of key contributions to research on moral development, conflict, and children's close relationships. Our findings build on past theory and research delineating how children's construals and judgments of conflict experiences are embedded in their differentiated histories of interactions with particular others (e.g., Slomkowski & Killen, 1992; Wainryb & Turiel, 1994) and documenting the manner in which children's own everyday experiences of harming others may contribute to the construction of moral agency (Pasupathi & Wainryb, 2010a). More specifically, our results reveal how, as they reflect on their own perpetration of harm against siblings and friends, children may struggle with the varied ways in which their own intentions, desires, and interpretations may inevitably clash with those of others. In this respect, our results provide new insight into how these relationships may make distinct contributions to children's understandings of themselves and others as imperfect but fundamentally moral agents. Finally, our results may have implications for how parents and educators can intervene most fruitfully into children's conflicts with siblings and friends; by taking children's varied interpretations of their own experiences as a starting point (e.g., their relative emphasis on their own vs. others' perspectives), adults may be able to more effectively scaffold children's moral understandings across a wide variety of relationships.

References

Arsenio, W., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen, & J. Smetana (Eds.), *Handbook of moral development* (pp. 581–610). Mahwah, NJ: Erlbaum.

Astor, R. A. (1994). Children's moral reasoning about family and peer violence: The role of provocation and retribution. *Child Development*, *65*, 1054–1067. doi:10.2307/1131304

Bank, S. P., & Kahn, M. D. (1982). *The sibling bond*. New York: Basic Books.

Buhrmester, D. (1992). The developmental courses of sibling and peer relationships. In F. Boer & J. Dunn (Eds.), *Children's sibling relationships: Developmental and clinical issues* (pp. 19–40). Hillsdale, NJ: Erlbaum.

Bukowski, W. M., & Sippola, L. K. (1996). Friendship and morality: (How) are they related? In W. M. Bukowski, A. F. Newcomb, & W. W. Hartup (Eds.), *The company they keep: Friendships in childhood and adolescence* (pp. 238–261). New York: Cambridge University Press.

Conger, K. J., & Little, W. M. (2010). Sibling relationships during the transition to adulthood. *Child Development Perspectives*, *4*, 87–94. doi:10.1111/j.1750-8606.2010.00123.x

Damon, W. (1977). *The social world of the child*. San Francisco, CA: Jossey-Bass.

DeHart, G. (1999). Conflict and averted conflict in preschoolers' interactions with siblings and friends. In W. A. Collins, & B. Laursen (Eds.), *Relationships in developmental contexts: The Minnesota symposia on child development* (Vol. 30, pp. 281–303). Mahwah, NJ: Erlbaum.

Dunn, J. (1984). *Sisters and brothers*. Cambridge, MA: Harvard University Press.

Dunn, J. (1993). *Young children's close relationships: Beyond attachment*. Newbury Park, CA: Sage.

Dunn, J., Slomkowski, C., Donelan, N., & Herrera, C. (1995). Conflict, understanding, and relationships: Developments and differences in the preschool years. *Early Education and Development*, *6*, 297–311. doi:10.1207/s15566935eed0604_2

Fonzi, A., Schneider, B. H., Tani, F., & Tomada, G. (1997). Predicting children's friendship status from their dyadic interaction in structured situations of potential conflict. *Child Development*, *68*, 496–506. doi:10.2307/1131674

Horn, S. S. (2006). Group status, group bias, and adolescents' reasoning about the treatment of others in school contexts. *International Journal of Behavioral Development*, *30*, 208–218. doi:10.1177/0165025406066721

Howe, N., & Recchia, H. E. (2008). *Siblings and sibling rivalry*. In M. M. Haith & J. B. Benson (Eds.), *Encyclopedia of infant and early childhood development* (Vol. 3, pp. 154–164). San Diego, CA: Academic Press.

Hughes, C., & Dunn, J. (2002). "When I say a naughty word." A longitudinal study of young children's accounts of anger and sadness in themselves and close others. *British Journal of Developmental Psychology*, *20*, 515–535. doi:10.1348/026151002760390837

Kim, J.-Y., McHale, S. M., Osgood, D. W., & Crouter, A. C. (2006). Longitudinal course and family correlates of sibling relationships from childhood through adolescence. *Child Development*, *77*, 1746–1761. doi:10.1111/j.1467-8624.2006.00971.x

Komolova, M., & Wainryb, C. (2011). "What I want and what you want": Children's thinking about competing personal preferences. *Social Development*, *20*, 334–352. doi:10.1111/j.1467-9507.2010.00589.x

Laursen, B., & Collins, W. A. (1994). Interpersonal conflict during adolescence. *Psychological Bulletin*, *115*, 197–209. doi:10.1037/0033-2909.115.2.197

Laursen, B., Finkelstein, B. D., & Betts, N. D. (2001). A developmental meta-analysis of peer conflict resolution. *Developmental Review*, *21*, 423–499. doi:10.1006/drev.2000.0531

MacEvoy, J. P., & Asher, S. R. (2012). When friends disappoint: Boys' and girls' responses to transgressions of friendship expectations. *Child Development*, *83*, 104–119. doi:10.1111/j.1467-8624.2011.01685.x

Pasupathi, M., & Wainryb, C. (2010a). Developing moral agency through narrative. *Human Development*, *53*, 55–80. doi:10.1159/000288208

Pasupathi, M., & Wainryb, C. (2010b). On telling the whole story: Facts and interpretations in autobiographical memory narratives from childhood through midadolescence. *Developmental Psychology*, *46*, 735–746. doi:10.1037/a0018897

Patterson, G. R. (1986). The contribution of siblings to training for fighting: A microsocial analysis. In D. Olweus, J. Block, & M. Radke-Yarrow (Eds.), *Development of antisocial and prosocial behavior: Research, theories, and issues* (pp. 235–261). New York: Academic Press.

Piaget, J. (1932). *The moral judgment of the child*. New York: Harcourt Brace.

Raffaelli, M. (1997). Young adolescents' conflicts with siblings and friends. *Journal of Youth and Adolescence*, *26*, 539–558. doi:10.1023/A:1024529921987

Recchia, H. E., & Howe, N. (2010). When do siblings compromise? Associations with children's descriptions of conflict issues, culpability, and emotions. *Social Development*, *19*, 838–857. doi:10.1111/j.1467-9507.2009.00567.x

Recchia, H. E., Wainryb, C., & Pasupathi, M. (2011, April). Helping and hurting: The construction of moral agency in children's conversations with their mothers. In C. Wainryb & H. E. Recchia (Chairs), *Parent-child conversations as contexts for moral development*. Symposium conducted at the biennial conference of the Society for Research in Child Development, Montreal, Canada.

Reis, H. T., Collins, W. A., & Berscheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, *126*, 844–872. doi:10.1037/0033-2909.126.6.844

Ross, H. S., Filyer, R. E., Lollis, S. P., Perlman, M., & Martin, J. P. (1994). Administering justice in the family. *Journal of Family Psychology*, *8*, 254–273. doi:10.1037/0893-3200.8.3.254

Ross, H. S., Siddiqui, A., Ram, A., & Ward, L. (2004). Perspectives on self and other in children's representations of sibling conflict. *International Journal of Behavioral Development*, *28*, 37–47. doi:10.1080/01650250344000253

Selman, R. (1980). *The growth of interpersonal understanding*. New York: Academic Press.

Shantz, D. W., & Pentz, T. (1972). Situational effects on justifiableness of aggression at three age levels. *Child Development*, *43*, 274–281. doi:10.2307/1127893

Slomkowski, C. L., & Killen, M. (1992). Young children's conceptions of transgressions with friends and nonfriends. *International Journal of Behavioral Development*, *15*, 247–258.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Erlbaum.

Smetana, J. G., & Asquith, P. (1994). Adolescents' and parents' conceptions of parental authority and personal autonomy. *Child Development*, *65*, 1147–1162. doi:10.2307/1131311

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: Norton.

Thompson, R. A. (2006). The development of the person: Social understanding, relationships, self, conscience. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg, (Vol. Ed.), *Handbook of child psychology*: *Vol. 3. Social, emotional, and personality development* (6th ed., pp. 24–98). New York: Wiley.

Updegraff, K. A., McHale, S. M., & Crouter, A. C. (2002). Adolescents' sibling relationship and friendship experiences: Developmental patterns and relationship linkages. *Social Development*, *11*, 182–204. doi:10.1111/1467-9507.00194

Vandell, D. L., & Bailey, M. D. (1992). Conflicts between siblings. In C. U. Shantz, & W. W. Hartup (Eds.), *Conflict in child and adolescent development* (pp. 242–269). Cambridge, England: Cambridge University Press.

Vespo, J. E., & Caplan, M. (1993). Preschoolers' differential conflict behavior with friends and acquaintances.

Early Education and Development, *4*, 45–53. doi:10.1207/s15566935eed0401_4

Wainryb, C., Brehl, B., & Matwin, S. (2005). Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal conflicts. *Monographs of the Society for Research in Child Development*, *70* (Serial No. 281). doi:10.1111/j.1540-5834.2005.00350.x

Wainryb, C., Shaw, L. A., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology*, *37*, 373–386. doi:10.1037/0012-1649.37.3.373

Wainryb, C., & Turiel, E. (1994). Dominance, subordination, and concepts of personal entitlements in cultural contexts. *Child Development*, *65*, 1701–1722. doi:10.2307/1131289

Walton, M. D., & Brewer, C. L. (2001). The role of personal narrative in bringing children into the moral discourse of their culture. *Narrative Inquiry*, *11*, 307–334. doi:10.1075/ni.11.2.04wal

Wilson, A. E., Smith, M. D., Ross, H. S., & Ross, M. (2004). Young children's personal accounts of their sibling disputes. *Merrill-Palmer Quarterly*, *50*, 39–60. doi:10.1353/mpq.2004.0008

Youniss, J. (1980). *Parents and peers in social development: A Sullivan-Piaget perspective*. Chicago: University of Chicago Press.

Youniss, J., & Smollar, J. (1985). *Adolescent relations with mothers, fathers, and friends*. Chicago: University of Chicago Press.

Child Development, July/August 2013, Volume 84, Number 4, Pages 1475–1490

Predicting Long-Term Growth in Students' Mathematics Achievement: The Unique Contributions of Motivation and Cognitive Strategies

Kou Murayama, Reinhard Pekrun,
and Stephanie Lichtenfeld
University of Munich

Rudolf vom Hofe
University of Bielefeld

This research examined how motivation (perceived control, intrinsic motivation, and extrinsic motivation), cognitive learning strategies (deep and surface strategies), and intelligence jointly predict long-term growth in students' mathematics achievement over 5 years. Using longitudinal data from six annual waves (Grades 5 through 10; $M_{\text{age}} = 11.7$ years at baseline; $N = 3{,}530$), latent growth curve modeling was employed to analyze growth in achievement. Results showed that the initial level of achievement was strongly related to intelligence, with motivation and cognitive strategies explaining additional variance. In contrast, intelligence had no relation with the growth of achievement over years, whereas motivation and learning strategies were predictors of growth. These findings highlight the importance of motivation and learning strategies in facilitating adolescents' development of mathematical competencies.

For decades, researchers in developmental and educational psychology have been concerned with the determinants of academic achievement. Today, there seems to be agreement that both motivational and strategy variables play an important role in explaining academic achievement (for a review, see Robbins et al., 2004). However, although a large portion of previous research has focused on the relation between these variables and academic achievement assessed at a particular time point, fewer studies have investigated whether motivational and strategy variables predict long-term *growth* in academic achievement. This is unfortunate because one of the ultimate goals in education is to facilitate sustainable learning (long-term intraindividual growth, i.e., change relative to current achievement of the individual), rather than to focus on performance attainment at one point in time. Clearly, systematic investigation of the long-term determinants of growth in academic achievement is imperative. In the present research, using longitudinal data from six annual waves over the adolescent years, we examined a variety of motivational variables and cognitive strategies as predictors of both concurrent level and long-term growth in math achievement.

Motivation: Perceived Control and Intrinsic–Extrinsic Motivation

Past studies have identified various motivational factors that affect academic achievement. The present research focused on three of these factors that are regarded as especially important in motivation theory and research: perceived control, intrinsic motivation, and extrinsic motivation (Eccles & Wigfield, 2002; Pekrun, 2006). *Perceived control* is conceptualized as subjective appraisal of the causal link between one's action and outcomes (Perry, Hladkyj, Pekrun, & Pelletier, 2001; Rotter, 1966). That is, perceived control reflects one's expectancy to obtain a desired outcome through an action. Concepts of perceived control and competence (and related constructs of expectancies) play a prominent role in motivation theories (Bandura, 1977; Eccles & Wigfield, 2002; Marsh & Shavelson, 1985; Pekrun, 1993), and many studies have shown that perceived control and competence in academic domains are positively related to achievement (Marsh, 1990; Meece, Wigfield, & Eccles, 1990). By implication, it is straightforward to expect that perceived control

---

This study was supported by an Alexander von Humboldt Foundation fellowship (to K. Murayama) and four grants from the German Research Foundation (DFG; to R. Pekrun, Project for the Analysis of Learning and Achievement in Mathematics, PALMA; PE 320/11-1, PE 320/11-2, PE 320/11-3, PE 320/11-4).

Correspondence concerning this article should be addressed to Kou Murayama, Department of Psychology, University of Munich, Leopoldstrasse 13 (PF 67), Munich, Germany 80802. Electronic mail may be sent to murakou@orion.ocn.ne.jp.

© 2012 The Authors
Child Development © 2012 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2013/8404-0025
DOI: 10.1111/cdev.12036

is positively associated with concurrent level of students' achievement in domains such as mathematics.

Importantly, adolescents' perceived control has also been shown to relate to their learning. Perceived control is linked to active and effortful commitment to learning (Skinner, Wellborn, & Connell, 1990), persistence when performing difficult and challenging tasks (Cervone & Peake, 1986), and intrinsic engagement (Gottfried, 1990). Together these findings suggest that perceived control should help adolescent students acquire new knowledge, and as such, should positively predict growth in their achievement. Indeed, longitudinal research on students' academic self-concepts has produced robust evidence that competence-related self-appraisals can influence subsequent academic achievement (e.g., Marsh, 1990; for reviews, see Marsh & Craven, 2006; Valentine, DuBois, & Cooper, 2004). Although (past) self-concept research relied on lagged analyses that do not consider absolute growth over time, the findings suggest that competence-related appraisals such as perceived control can be a positive predictor of long-term change in students' academic achievement.

Intrinsic motivation is defined as motivation to engage in a task for the sake of interest in the task itself and the inherent pleasure and satisfaction derived from the task, whereas *extrinsic motivation* is defined as motivation to engage in a task for external reasons (Deci & Ryan, 1985). Extrinsic motivation is heterogeneous in that there are a variety of different external rewards that produce motivated behavior. In the present research, we focus on extrinsic motivation driven by the desire to get good grades, as this is one of the most prevalent forms of extrinsic motivation in academic settings (Lemos, 1996; Pekrun, 1993).

Intrinsic motivation is associated with various variables supporting learning, such as active and effortful engagement (Ryan & Connell, 1989), persistence in the face of failure (Elliott & Dweck, 1988), and positive emotional learning experiences (Pekrun, Goetz, Titz, & Perry, 2002). Extrinsic motivation, on the other hand, is typically driven by expected short-term benefits of learning and is linked to instrumental learning independent of interest (Grolnick & Ryan, 1987), overuse of dependent help seeking (Butler, 1998), and self-handicapping (Urdan & Midgley, 2001). As such, extrinsic motivation seems suited to benefit immediate, but rather ephemeral, academic achievement, whereas intrinsic motivation seems ideally suited to benefit enduring, long-term learning. Thus, we propose that intrinsic motivation positively predicts growth in academic achievement (and possibly concurrent achievement as well; see also Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005), whereas extrinsic motivation has short-term effects that are manifested in a link to current achievement, but does not promote long-term growth. Although these time-dependent relations (i.e., relations that emerge at a particular point in time) of intrinsic and extrinsic motivation are documented in a few experimental studies (Murayama & Elliot, 2011; Vansteenkiste, Simons, Lens, Soenens, & Matos, 2005), little research has demonstrated such a pattern of effects in the context of long-term growth of academic achievement.

Cognitive Strategies: Deep and Surface Learning Strategies

Learning strategies are strategies that are employed by learners to study materials. Herein, we focus on two primary cognitive learning strategies that are widely recognized as having conceptual and predictive utility: *deep learning strategies* (specifically, elaboration of learning material) and *surface learning strategies* (rehearsal or memorization; Ramsden, 1988). Deep strategies involve challenging the veracity of information encountered and attempting to integrate new information with prior knowledge, whereas surface learning strategies characterize the repetitive rehearsal and rote memorization of information (Entwistle & Ramsden, 1983).

The extant data are mixed for the relation between these learning strategies and academic achievement, but tend to show null or positive relations for deep strategies and null or negative relations for surface strategies (Entwistle & Ramsden, 1983; Meece, Blumenfeld, & Hoyle, 1988). However, most of the studies to date have focused on links with the level of academic achievement at one point in time. A few studies have shown that learning-related strategy use has an impact on the growth of academic achievement during early childhood (e.g., McClelland, Acock, & Morrison, 2006), but the long-term relations of learning strategies with achievement during adolescence remain unclear.

Deep learning involves semantic understanding of study materials. Semantic understanding is an essential component in acquiring durable and meaningful knowledge, and the knowledge acquired through this semantic elaboration is more likely to lead to later academic success. Therefore, we anticipated that deep strategies would have positive relations with growth (as well as initial levels) of academic achievement. In contrast, we expected that surface strategies would not predict growth rates in achievement. Surface strategies involve rote memorizing without deep elaboration, and studies have indicated that knowledge acquired through rote memorizing fades quickly (Brown & Craik, 2000). Therefore, we anticipated that surface learning would not support further learning, although these strategies might help immediate academic achievement.

Limitations in Previous Research on Long-Term Growth in Achievement

This study used latent growth curve modeling (McArdle, Anderson, Birren, & Schaie, 1990) to examine predictors of growth in math achievement. One of the advantages of latent growth curve modeling is that it can evaluate the predictors of absolute levels of growth. As noted earlier, previous theorizing and evidence suggest that some motivational variables and learning strategies should predict long-term growth in academic achievement. However, only a few studies have utilized latent growth curve modeling to directly address the predictors of absolute levels of growth in achievement over time. Furthermore, the research available that used latent growth curve modeling has a number of important limitations that substantially reduce possibilities to draw firm inferences.

First, there is a significant lack of work that controlled for intelligence. Intelligence is a ubiquitous, strong predictor of academic achievement (Neisser et al., 1996), implying that it may be the most important confounding variable when examining links between variables of learning, such as motivation and cognitive strategies, and academic achievement (Steinmayr & Spinath, 2009). It is easy to imagine students who have higher levels of intelligence to be more likely to be motivated and to use efficient strategies that result in their higher academic achievement. Therefore, it is important to include intelligence as a predictor to evaluate the predictive utility of motivation and cognitive strategies independently of basic cognitive abilities.

It should also be noted that the utility of intelligence to predict growth of achievement is by itself of theoretical importance. Although researchers agree that intelligence is a strong predictor of students' *level* of academic achievement, they have not reached consensus about its utility in predicting *growth* of achievement over time (Lohman, Ackerman, Kyllonen, & Roberts, 1999). Studies indicated that intelligence scores did not (or only weakly) predict growth curves in performance attainment (e.g., Gutman, Sameroff, & Cole, 2003; Underwood, Boruch, & Malmi, 1978; Woodrow, 1946). These findings suggest that, unlike motivation and cognitive strategies, intelligence may not reflect the capacity to learn new knowledge and skills. Importantly, however, the majority of these studies focused on short-term learning (sometimes learning within one experimental session). Studies investigating the relations of intelligence with the long-term growth of academic achievement are still rare.

Second, much of the previous literature using latent growth curve modeling failed to establish a common metric for assessments of achievement over time. For analyzing growth in academic achievement, if one simply uses achievement test scores or teacher-provided grades, it is not possible to adequately study growth because these scores are likely not comparable across years. To establish a common metric, item response theory (IRT) can be employed, and anchor items can be included across adjacent time points to allow for vertical scaling (McDonald, 1999). Although IRT scaling is generally acknowledged today as an appropriate way to examine academic achievement over time, there still are many studies that have not used this methodology, casting doubt on the validity of their findings on growth (e.g., Gutman et al., 2003; Hart, Hofmann, Edelstein, & Keller, 1997; Johnson, McGue, & Iacono, 2006).

Third, previous studies included relatively small numbers of variables, focusing on either motivational or strategy constructs (e.g., McClelland et al., 2006; Shim, Ryan, & Anderson, 2008). However, given that motivational and strategy variables often show substantial correlations, it is important to include multiple variables from both groups of constructs to examine their unique contributions.

Finally, previous studies failed to consider the possibility that the utility of motivational and strategy variables for predicting academic achievement may change depending on students' developmental stage. Many studies have documented developmental trajectories of motivation and strategies over time (e.g., Fredricks & Eccles, 2002; Frenzel, Goetz, Pekrun, & Watt, 2010; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). It is possible, then, that not only the developmental trajectories but also the *functions* of motivation and learning strategies in shaping students' achievement during early adolescence can change over time. However, few studies have examined developmental change in the functions of these variables.

Overview of the Present Research and Hypotheses

The current study aimed to broaden the extant knowledge about how motivation and cognitive strategies relate to growth in academic achievement, using German longitudinal mathematics achievement data from six annual waves (Grades 5 through 10). We controlled for intelligence to evaluate the contributions of motivation and strategy variables, and utilized IRT to scale academic achievement over time. We included all of the motivation variables (perceived control, intrinsic motivation, and extrinsic motivation) and learning strategy variables (deep and surface strategies) together to investigate their unique contributions. In addition, we compared the predictive power of motivation and strategies at two different time points (Grades 5 and 7) to examine possible functional change in these variables. Grades 5 and 7 were chosen because they represent critical periods in the educational careers of the German students who participated in the study. Specifically, these students attended public schools in the German state of Bavaria that use ability tracking starting in Grade 5. Students are assigned to low-ability schools (*Hauptschule*), medium-ability schools (*Realschule*), and high-ability schools (*Gymnasium*). In Grade 7, students' assignments to low- versus medium-ability schools were reconsidered; thus, part of the student sample were reassigned to a different track. As such, for students in this study, Grades 5 and 7 represented critical career and academic transitions. Furthermore, students in Grades 5 and 7 are in their early adolescence, a period characterized by a great deal of intellectual, social, physical, and emotional change (Lerner, 1993), which may imply that the functions of academic motivation and strategies also undergo change.

Succinctly stated, the study tested the following hypotheses:

> Hypothesis 1: Perceived control is a positive predictor of the initial level as well as the growth rate of achievement.
>
> Hypothesis 2: Extrinsic motivation is a positive predictor of the initial level of achievement, whereas intrinsic motivation is a positive predictor of both the initial level and the growth rate of achievement.
>
> Hypothesis 3: Surface learning strategies are a positive predictor of the initial level of achievement, whereas deep learning strategies are a positive predictor of both the initial level and the growth rate of achievement.
>
> Hypothesis 4: Intelligence is a positive predictor of the initial level of achievement, but has no relation with growth rate.

Hypotheses 1–3 were tested while controlling for students' intelligence. Furthermore, although we did not have a specific hypothesis, we expected that the predictive power of motivation and strategies may change from Grade 5 to Grade 7.

Method

Participants and Design

The sample consisted of German students who participated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; see Frenzel et al., 2010; Frenzel, Pekrun, Dicke, & Goetz, in press; Pekrun et al., 2007). This project included a longitudinal study involving annual assessments during the secondary school years (Grades 5–10) to investigate the development of mathematics achievement. At each grade level, the PALMA math achievement test was administered toward the end of the school year. Students' intelligence and the self-reported motivational and strategy variables used in the current study were assessed toward the end of the school year in Grades 5 and 7.

Samples were drawn from secondary schools in the state of Bavaria, and were drawn so that they were sufficiently representative of the student population of Bavaria (Pekrun et al., 2007). The samples included students from all three school types within the Bavarian public school system as described earlier, including lower-track schools (*Hauptschule*), intermediate-track schools (*Realschule*), and higher-track schools (*Gymnasium*). These three school types differ in academic demands and students' entry-level academic ability. German students typically enter one of these schools in Grade 5 (i.e., the 1st year of the assessment), based largely on their prior academic achievement.

At the first assessment (Grade 5), the sample comprised 2,070 students from 42 schools (49.6% female, $M_{\text{age}} = 11.7$ years; 37.2% lower-track school students, 27.1% intermediate-track school students, and 35.7% higher-track school students). In each subsequent year, the study tracked the students who had participated in the previous assessment(s) and also included those students who had not yet participated in the study, but had become students of PALMA classrooms at the time of the assessment (see Pekrun et al., 2007). Overall, the study sample consisted of 3,530 students (49.7% female) who participated in at least one assessment (see Supporting Information Appendix S1 available online for further details).

Measures

Socioeconomic Status (SES)

Family SES was assessed by parent report using the EGP classification (Erikson, Goldthorpe, & Portocarero, 1979) which consists of six ordered categories of parental occupational status. In the current analysis, the scores are coded so that higher values represent higher family SES.

Mathematics Achievement

Mathematics achievement was assessed by the PALMA Mathematics Achievement Test (vom Hofe, Kleine, Pekrun, & Blum, 2005; vom Hofe, Pekrun, Kleine, & Götz, 2002). Using both multiple-choice and open-ended items, this test measures students' modeling competencies and algorithmic competencies in arithmetics, algebra, and geometry. It also comprises subscales pertaining to more specific contents (e.g., fractions, ratios, or functions).

The test was constructed using multimatrix sampling with a balanced incomplete block design (for details, see vom Hofe et al., 2002). Specifically, for each measurement point, two different test versions were prepared that consisted of approximately 60–90 items each, and students completed one of these two test booklets. Anchor items were included to allow for the linkage of the two different test forms as well as the six different measurement points. The obtained achievement scores were scaled using one-parameter logistic item-response theory (Rasch scaling; Wu, Adams, Wilson, & Haldane, 2007), with $M = 100$ and $SD = 15$ at Grade 5 (i.e., the first measurement point). Additional analyses confirmed the unidimensionality and longitudinal invariance of the test scales (see Supporting Information Appendix S1 available online).

Motivation and Learning Strategies

Motivation and learning strategies were assessed by self-report scales developed for PALMA. All scales pertained to the domain of mathematics, and reliability and validity of all the scales were ascertained in two large-sample pilot studies (*N*s = 784 and 1,613; Pekrun et al., 2007). Coefficient Omega was used to estimate the reliability of the scale scores (see Table 1). Omega provides more fine-tuned estimates of scale reliability as compared with the commonly used Cronbach's Alpha (McDonald, 1999).

Table 1
Descriptive Statistics and Internal Consistencies of the Predictors

	M	*SD*	Observed range	Coefficient of reliability
Grade 5				
Perceived control	3.87	0.72	1.17–5.00	.75
Intrinsic motivation	3.21	1.18	1.00–5.00	.87
Extrinsic motivation	3.37	0.99	1.00–5.00	.80
Deep learning strategies	2.38	0.78	1.00–4.00	.67
Surface learning strategies	2.49	0.75	1.00–4.00	.64
Intelligence	104.75	13.41	61–132	—
Grade 7				
Perceived control	3.43	0.83	1.00–5.00	.81
Intrinsic motivation	2.57	1.07	1.00–5.00	.84
Extrinsic motivation	2.87	0.97	1.00–5.00	.78
Deep learning strategies	2.03	0.70	1.00–4.00	.64
Surface learning strategies	2.26	0.73	1.00–4.00	.65
Intelligence	99.61	14.56	55–145	—

Perceived control. Pekrun et al.'s (2007) Perceived Academic Control scale was used to assess students' perceived control in mathematics (six items; e.g., "When doing math, the harder I try, the better I perform"). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale.

Intrinsic motivation. Three items from Pekrun's (1993) Intrinsic Academic Motivation scale were used to assess students' intrinsic motivation in mathematics (e.g., "I invest a lot of effort in math, because I am interested in the subject"). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale.

Extrinsic motivation. Four items from Pekrun's (1993) Extrinsic Academic Motivation scale were used to assess students' extrinsic motivation in mathematics (sample item: "In math I work hard, because I want to get good grades"). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale.

Deep learning strategies. The PALMA elaboration strategies scale (Pekrun et al., 2007) was used to assess students' deep learning strategies (three items; e.g., "When I study for exams, I try to make connections with other areas of math"). Participants responded on a 1 (*strongly disagree*) to 4 (*strongly agree*) scale.

Surface learning strategies. The PALMA rehearsal strategies scale (Pekrun et al., 2007) was used to assess students' surface learning strategies (three items; sample item: "For some math problems I memorize the steps to the correct solution"). Participants responded on a 1 (*strongly disagree*) to 4 (*strongly agree*) scale.

Intelligence. Intelligence was measured using the 25-item nonverbal reasoning subtest of the German adaptation of Thorndike's Cognitive Abilities Test (Kognitiver Fähigkeitstest [KFT 4–12 + R]; Heller & Perleth, 2000).

Strategy of Data Analysis

Growth curve models can be fit in both structural equation modeling (SEM) and multilevel modeling frameworks. We decided to use the SEM framework because it allows us to assess the overall fit between the empirical data and the growth curve models we specified. Specifically, we evaluated the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). It should be noted that nearly all current SEM packages compute CFI and TLI based on an inappropriate baseline model when evaluating the fit of latent growth curve models (Wu, West & Taylor, 2009). In our analysis of latent growth curve models, we corrected the CFI and TLI by specifying the appropriate baseline model (see Wu et al., 2009). Following conventions, we regarded values higher than .90 for CFI and TLI, lower than .08 for the RMSEA, and lower than .10 for the SRMR as indicating an acceptable fit.

Prior to the analyses, all the predictors except for demographic variables were standardized with zero mean and unit variance, and SES was mean centered to facilitate the interpretation of the results. Based on Muthén and Muthén (2004), we adjusted the standard errors and chi-square statistics to correct for potential statistical biases resulting from non-normality of the data. Due to the longitudinal design of the study, there is a certain proportion of missing data. Accordingly, to make full use of the data from students who only partially participated in the investigation, we applied the full information maximum likelihood method to deal with missing data (Enders, 2006). Additional analyses that addressed issues of measurement error and the nested structure of the data (i.e., individuals nested within schools) are reported in Supporting Information Appendix S1 available online.

Results

Longitudinal Analysis of Measurement Invariance for Motivation and Strategies

The descriptive statistics of the predictor variables for Grades 5 and 7 are provided in Table 1. Because we wanted to compare the predictive utility of motivational and strategy variables across two different time points (i.e., Grades 5 and 7), we used confirmatory factor analysis to assess measurement equivalence across these time points. Specifically, we evaluated the full scalar and error variance invariance model (see Steenkamp & Baumgartner, 1998), in which the item intercept, factor loadings, factor variances and covariances, and error variances were set to be equal across the time points. Support for this model would provide strong evidence for longitudinal invariance of the measures. All of the motivation and learning strategy variables were analyzed together. The residuals associated with indicators of the same items in Grades 5 and 7 were allowed to correlate. The results supported the longitudinal invariance of the scales, with good fit to the data, $\chi^2(663) = 2{,}548.98$, $p < .01$, CFI = .92, TLI = .92, RMSEA = .031, SRMR = .060. These findings suggest that it is legitimate to compare the predictive utility of the motivational and strategy variables across time points.

The disattenuated correlations of the latent factors for the same constructs over time were .37 to .47 ($ps < .01$). The size of these coefficients indicates that the motivational and strategy variables were to some extent stable across the 2 years but also showed change, suggesting that it is possible that their functional roles for predicting academic achievement also changed over time. Comparison of factor means indicated a decline in all of the motivation and strategy variables from Grades 5 to 7 ($ps < .01$), which is largely consistent with previous findings (e.g., Jacobs et al., 2002; Otis, Grouzet & Pelletier, 2005).

Latent Growth Curve Modeling

In the latent growth curve analysis, we first constructed a growth curve model without motivational and strategy variables to specify the basic functional form of growth in mathematical achievement. In this model, math achievement scores across six different time points (Grades 5–10) are modeled by a set of common factors representing the functional form of growth. Our initial exploration of the data indicated that a linear growth model (a model including an intercept and a linear slope) did not fit the data well, suggesting that the growth curve includes nonlinear change. Accordingly, we decided to use an exponential growth curve model (Grimm, Ram, & Hamagami, 2011; for details about model selection, see Supporting Information Appendix S1 available online).

Exponential growth curve modeling has several advantages over the commonly used quadratic growth curve model. Notably, exponential growth models provide interpretable parameters describing growth, such as the total amount of change over time. This point is of particular importance for our study, as we are interested in the factors that influence the overall amount of growth over years. The model is represented by the following equation.

$$y_t = \eta_0 + \eta_1[1 - \exp(-\alpha \cdot x_t)] + \varepsilon_t; \quad t = 5, 6, 7, 8, 9, 10 \tag{1}$$

where y_t is the math achievement score at grade t, and η_0 and η_1 are the latent variables representing the intercept and total amount of (asymptotic) growth. α is a free parameter representing the rate of growth. ε_t is the error term for grade t, and we constrained the variances of this term to be equal across the time points [i.e., $\mathrm{var}(\varepsilon_5) = \mathrm{var}(\varepsilon_6) = \mathrm{var}(\varepsilon_7) = \mathrm{var}(\varepsilon_8) = \mathrm{var}(\varepsilon_9) = \mathrm{var}(\varepsilon_{10})$; Hertzog, von Oertzen, Ghisletta, & Lindenberger, 2008]. x_t represents the fixed time coding at grade t anchored at the initial time point [i.e., $x_5 = 0$, $x_6 = 1$, $x_7 = 2$, $x_8 = 3$, $x_9 = 4$, $x_{10} = 5$]. By this particular coding of time point, we can interpret η_0 as the predicted math achievement score at the initial time point, and η_1 as the asymptotic total amount of growth from the initial time point.

To control for participants' demographic background, variables representing gender, school type, and SES were included in the model. Gender was treated as a dummy-coded variable (GENDER; 0 = female, 1 = male), and school type was represented by two orthogonal contrast variables. The first school-type variable represents the difference between higher-track schools and the pooled intermediate- and lower-track schools (SCTYPE_1; higher-track school = 2, intermediate-track school = −1, lower-track school = −1). The second variable represents the difference between intermediate-track schools and lower-track schools (SCTYPE_2; higher-track school = 0, intermediate-track school = 1, lower-track school = −1). The following equations show the mathematical representation of this part of the model.

$$\begin{aligned} \eta_0 = \mu_{\eta 0} + \gamma_{01}\mathrm{GENDER} + \gamma_{02}\mathrm{SCTYPE}_1 \\ + \gamma_{03}\mathrm{SCTYPE}_2 + \gamma_{04}\mathrm{SES} + \zeta_0 \end{aligned} \tag{2}$$

$$\begin{aligned} \eta_1 = \mu_{\eta 1} + \gamma_{11}\mathrm{GENDER} + \gamma_{12}\mathrm{SCTYPE}_1 \\ + \gamma_{13}\mathrm{SCTYPE}_2 + \gamma_{14}\mathrm{SES} + \zeta_1 \end{aligned} \tag{3}$$

where γ represents the path coefficients of the corresponding variable. ζ is the error (or disturbance) term for each growth component.

The model showed an acceptable fit to the data, $\chi^2(36) = 757.48$, $p < .01$, CFI = .96, TLI = .94, RMSEA = .075, SRMR = .090. The results (see Table 2) revealed that the mean total amount of growth was significantly positive ($\mu_{\eta 1} = 144.17$, $p < .01$) with a positive growth rate ($\alpha = .06$, $p < .01$). In addition, most of the path coefficients for gender, school type, and SES variables were significant, suggesting variability in growth trajectories across demographic background variables. Figure 1 displays the estimated growth curves for male and female students from different types of schools.

An interesting observation in Figure 1 is that the difference between school tracks became larger with increasing grade level, as indicated by the significantly positive relation of school type with total amount of change ($\gamma_{12} = 12.74$, $p < .01$; $\gamma_{13} = 9.27$, $p < .01$). This finding suggests that the tracking system used in Germany may be a source of a "Matthew effect" in math achievement—a phenomenon where the gap between high-achieving and low-achieving students increases with time (Bast & Reitsma, 1997; Baumert, Nagy, & Lehmann, 2012). To quantitatively evaluate the link between tracking and the Matthew effect, we

Table 2
Parameter Estimates for the Growth Curve Model Including Gender and School Type

	Estimate	*SE*
Intercept		
Mean ($\mu_{\eta 0}$)	99.28**	0.31
Gender (γ_{01})	3.24**	0.45
School type 1 (high vs. mid or low track; γ_{02})	4.51**	0.17
School type 2 (mid vs. low track; γ_{03})	6.24**	0.27
Family socioeconomic status (γ_{03})	0.66**	0.13
Total amount of growth		
Mean ($\mu_{\eta 1}$)	144.17**	18.1
Gender (γ_{11})	−1.08	2.37
School type 1 (high vs. mid or low track; γ_{12})	9.27**	1.29
School type 2 (mid vs. low track; γ_{13})	12.74**	2.02
Family socioeconomic status (γ_{13})	0.85	0.67

$^{**}p < .01$.

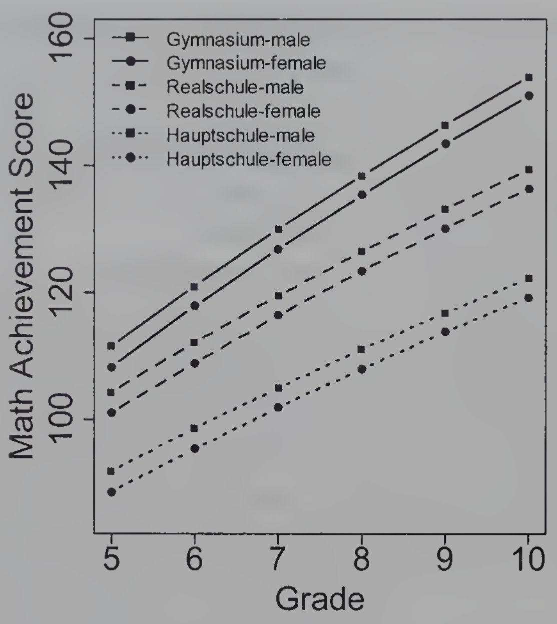


Figure 1. Estimated growth curve of math achievement from Grades 5 to 10 as a function of gender and school type.
Note. Gymnasium = higher-track school; Realschule = intermediate-track school; Hauptschule = lower-track school.

computed the correlation between the intercept (η_0) and the total amount of change (η_1) before and after including the school-type variables. When the school-type variables were not included, this correlation was positive ($r = .29$, $p < .01$), indicating the presence of a Matthew effect. Controlling for these variables dramatically decreased the correlation ($r = .01$, $p = .79$), suggesting that the Matthew effect observed in the data is linked to tracking as used in the German secondary school system.

Predictive Relations of Motivation and Strategies Assessed at Grade 5

To examine the relations of Grade 5 motivational and strategy variables with the growth curve of achievement, we included these variables as predictors of the growth components. We also included intelligence as a predictor to control for students' basic cognitive abilities and to investigate the predictive utility of intelligence for the growth components. Specifically, Equations 2 and 3 were expanded to take the following form:

$$\begin{aligned}\eta_0 =\ &\mu_{\eta 0} + \gamma_{01}\mathrm{GENDER} + \gamma_{02}\mathrm{SCTYPE}_1 \\ &+ \gamma_{03}\mathrm{SCTYPE}_2 + \gamma_{04}\mathrm{SES} + \gamma_{05}\mathrm{CONT} \\ &+ \gamma_{06}\mathrm{INT} + \gamma_{07}\mathrm{EXT} + \gamma_{08}\mathrm{DEEP} + \gamma_{09}\mathrm{SURFACE} \\ &+ \gamma_{10}\mathrm{IQ} + \zeta_0\end{aligned} \tag{4}$$

$$\begin{aligned}\eta_1 =\ &\mu_{\eta 1} + \gamma_{11}\mathrm{GENDER} + \gamma_{12}\mathrm{SCTYPE}_1 \\ &+ \gamma_{13}\mathrm{SCTYPE}_2 + \gamma_{14}\mathrm{SES} + \gamma_{15}\mathrm{CONT} \\ &+ \gamma_{16}\mathrm{INT} + \gamma_{17}\mathrm{EXT} + \gamma_{18}\mathrm{DEEP} \\ &+ \gamma_{19}\mathrm{SURFACE} + \gamma_{20}\mathrm{IQ} + \zeta_1\end{aligned} \tag{5}$$

where CONT, INT, EXT, DEEP, SURFACE, and IQ represent the variables of perceived control, intrinsic motivation, extrinsic motivation, deep learning strategies, surface learning strategies, and intelligence assessed at Grade 5.

The model showed an acceptable fit to the data, $\chi^2(60) = 843.00$, $p < .01$, CFI = .97, TLI = .93, RMSEA = .061, SRMR = .059. The parameter estimates of the model are presented in Table 3. Not surprisingly, intelligence had a strong positive relation with the intercept ($\gamma_{10} = 4.72$, $p < .01$). In addition, however, most of the motivational and

Table 3

Parameter Estimates for the Growth Curve Model Including Motivational and Strategy Variables Assessed at Grade 5

	Estimate	*SE*
Intercept		
Mean ($\mu_{\eta 0}$)	99.68**	0.30
Gender (γ_{01})	2.42**	0.42
School type 1 (high vs. mid or low track; γ_{02})	3.36**	0.17
School type 2 (mid vs. low track; γ_{03})	4.67**	0.27
Family socioeconomic status (γ_{04})	0.45**	0.12
Perceived control (γ_{05})	2.21**	0.27
Intrinsic motivation (γ_{06})	1.53**	0.29
Extrinsic motivation (γ_{07})	0.74**	0.28
Deep learning strategies (γ_{08})	−0.41	0.31
Surface learning strategies (γ_{09})	−1.60**	0.27
Intelligence (γ_{10})	4.72**	0.27
Total amount of change		
Mean ($\mu_{\eta 1}$)	142.34**	17.6
Gender (γ_{11})	−1.22	2.49
School type 1 (high vs. mid or low track; γ_{12})	8.15**	1.26
School type 2 (mid vs. low track; γ_{13})	11.69**	2.00
Family socioeconomic status (γ_{14})	0.66	0.66
Perceived control (γ_{15})	3.78*	1.72
Intrinsic motivation (γ_{16})	1.04	1.84
Extrinsic motivation (γ_{17})	1.56	1.72
Deep learning strategies (γ_{18})	−2.52	1.88
Surface learning strategies (γ_{19})	−5.99**	1.71
Intelligence (γ_{20})	0.24	1.61

*$p < .05$. **$p < .01$.

strategy factors yielded significant incremental relations with the initial achievement score, over and above intelligence. Specifically, perceived control, intrinsic motivation, and extrinsic motivation were significant predictors of the intercept factor ($\gamma_{05} = 2.21$, $\gamma_{06} = 1.53$, and $\gamma_{07} = 0.74$, $ps < .01$), whereas surface learning strategies were a negative predictor ($\gamma_{09} = -1.60$, $p < .01$). Deep learning strategies did not show any significant relation ($p = .18$).

Intriguingly, in contrast to the relations with the intercept factor, intelligence did not predict the total amount of growth of math achievement over time ($\gamma_{20} = 0.24$, $p = .88$). However, two of the motivational and strategy variables turned out to be significant predictors of growth of math achievement over time. Specifically, perceived control was a significantly positive predictor of the slope factor ($\gamma_{15} = 3.78$, $p < .05$), whereas surface learning strategies were a significantly negative predictor ($\gamma_{19} = -5.99$, $p < .01$). This result underscores the unique importance of motivational and strategy variables in affecting not only current achievement but also the long-term growth of achievement.

To facilitate the interpretation of the findings on growth, we estimated the growth curves of students with scores of 1.5 *SD* above the mean for perceived control and 1.5 *SD* below the mean for surface learning strategies ("high-growth students"), and students with scores of 1.5 *SD* below the mean for perceived control and 1.5 *SD* above the mean for surface learning strategies ("low-growth students"). To avoid redundancy, Figure 2 presents the estimated growth curves only for male higher-track (Gymnasium) students. The shape of the growth curves for females and for intermediate- or lower-track students was basically the same. Figure 2 visually confirms that the slope is steeper for high-growth than for low-growth students.

Predictive Relations of Motivation and Strategies Assessed at Grade 7

The previous analysis partially supported our hypotheses in that (a) intelligence had strong positive relations with initial level of math achievement but showed null relations with growth rate and (b) perceived control had positive relations not only with initial level but also with the growth rate. To further explore our hypotheses, we next focused on students' growth in math achievement from Grades 7–10. Specifically, we investigated whether the growth curve estimated in the Grades 7–10 data can be predicted by motivation and learning strategies assessed at the Grade 7. The model was the same as in the previous analysis, Equations 1, 4, and 5, with the following exceptions. First, Equations 4 and 5 included the motivational and strategy variables and intelligence assessed at Grade 7 instead of Grade 5, as our focus was to examine the predictive utility of the variables assessed at Grade 7. Participants' school type was also coded based on the information at Grade 7. Second, we used the math achievement scores assessed across Grades 7–10, that is, $t = 7, 8, 9, 10$ in Equation 1 and $x_7 = 0$, $x_8 = 1$, $x_9 = 2$, $x_{10} = 3$. Third, we fixed the growth rate parameter to the value obtained in the previous analysis (i.e., $\alpha = .06$) to make the Grade 5 and Grade 7 analyses comparable. This constraint improved model fit in terms of the Akaike information criterion and the Bayesian information criterion, and did not significantly increase the chi-square fit statistics, $\Delta\chi^2(1) = 1.09$, $p = .30$.

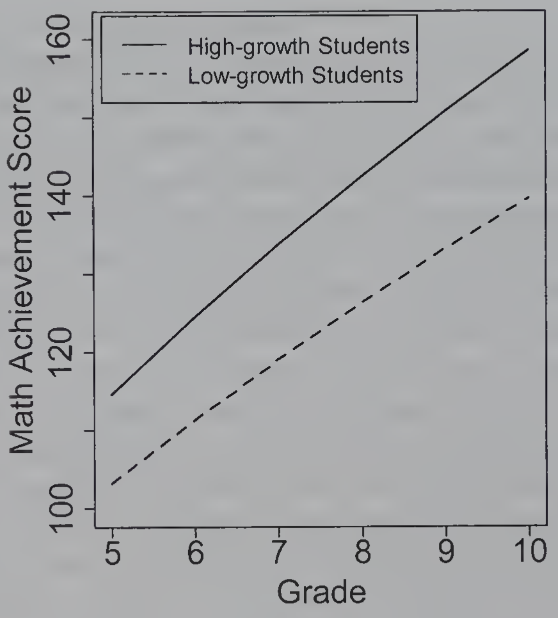


Figure 2. Estimated math achievement growth curve from 5th to 10th grades for high-growth and low-growth students.
Note. To avoid redundancy, only male Gymnasium students are plotted. High-growth students = 1.5 *SD* above the mean for perceived control and 1.5 *SD* below the mean for surface learning strategies. Low-growth students = 1.5 *SD* below the mean for perceived control and 1.5 *SD* above the mean for surface learning strategies.

The model showed an acceptable fit to the data, $\chi^2(28) = 477.79$, $p < .01$, CFI = .97, TLI = .91, RMSEA = .073, SRMR = .060. The parameter estimates of the model are presented in Table 4. For the prediction of the intercept, the pattern of results was almost the same as in the Grade 5 analysis; in addition to the strong positive explanatory power of intelligence ($\gamma_{10} = 6.10$, $p < .01$), perceived control and extrinsic motivation were significantly positive predictors of the initial level of math achievement ($\gamma_{05} = 3.35$, $p < .01$; $\gamma_{07} = 1.18$, $p < .01$), whereas surface learning strategies were a significantly negative predictor

($\gamma_{09} = -1.49$, $p < .01$). These results again corroborate the importance of motivational and strategy factors in accounting for current math achievement over and above intelligence. On the other hand, intrinsic motivation and deep learning strategies were not significant predictors of the intercept factor.

Intriguingly, for the prediction of total amount of growth, the Grade 7 analysis showed a pattern of results that differed substantially from the Grade 5 findings. Specifically, whereas perceived control, extrinsic motivation, and surface learning strategies did not predict growth of math achievement, intrinsic motivation and deep learning strategies were significantly positive predictors of the total amount of growth ($\gamma_{16} = 4.51$, $p < .05$; $\gamma_{18} = 4.64$, $p < .05$). Again, intelligence had null relations with the amount of growth ($\gamma_{20} = 0.37$, $p = .87$).

As in the Grade 5 analysis, these findings indicate that motivational and strategy variables uniquely contribute to growth of math achievement. Furthermore, these findings also indicate that there is developmental change in the predictive power of these variables. As assessed at Grade 7, intrinsic motivation and deep learning strategies did not have any significant relations with the intercept factor but related to subsequent growth, suggesting that they may not have immediate utility but may provide long-term benefits in facilitating math achievement at this age. As an illustration for the nature of this finding, Figure 3 displays the estimated growth curves of male higher-track (Gymnasium) students with scores 1.5 *SD* above the mean for intrinsic motivation and deep learning strategies ("high-growth students") and students with scores 1.5 *SD* below the mean for these variables ("low-growth students"). The shape of the growth curves for females and for intermediate- or lower-track students was basically the same. Figure 3 visually confirms that high-growth and low-growth students did not differ in terms of the initial status at Grade 7, but developed differently across the subsequent grade levels.

Table 4

Parameter Estimates for the Growth Curve Model Including Motivational and Strategy Variables Assessed at Grade 7

	Estimates	*SE*
Intercept		
Mean ($\mu_{\eta 0}$)	114.57**	0.28
Gender (γ_{01})	1.85**	0.43
School type 1 (high vs. mid or low track; γ_{02})	4.47**	0.17
School type 2 (mid vs. low track; γ_{03})	6.88**	0.28
Family socioeconomic status (γ_{04})	0.33**	0.12
Perceived control (γ_{05})	3.35**	0.27
Intrinsic motivation (γ_{06})	0.37	0.29
Extrinsic motivation (γ_{07})	1.18**	0.28
Deep learning strategies (γ_{08})	−0.49	0.28
Surface learning strategies (γ_{09})	−1.49**	0.25
Intelligence (γ_{10})	6.10**	0.27
Total amount of change		
Mean ($\mu_{\eta 1}$)	134.39**	2.41
Gender (γ_{11})	6.41	3.40
School type 1 (high vs. mid or low track; γ_{12})	3.80**	1.42
School type 2 (mid vs. low track; γ_{13})	6.23*	2.76
Family socioeconomic status (γ_{14})	1.69	0.96
Perceived control (γ_{15})	−2.76	2.18
Intrinsic motivation (γ_{16})	4.51*	2.25
Extrinsic motivation (γ_{17})	−0.55	2.10
Deep learning strategies (γ_{18})	4.64*	2.23
Surface learning strategies (γ_{19})	−0.81	2.02
Intelligence (γ_{20})	0.37	2.28

*$p < .05$. **$p < .01$.

Discussion

This research examined the predictive power of motivation, cognitive strategies, and intelligence for explaining the long-term growth of adolescents'

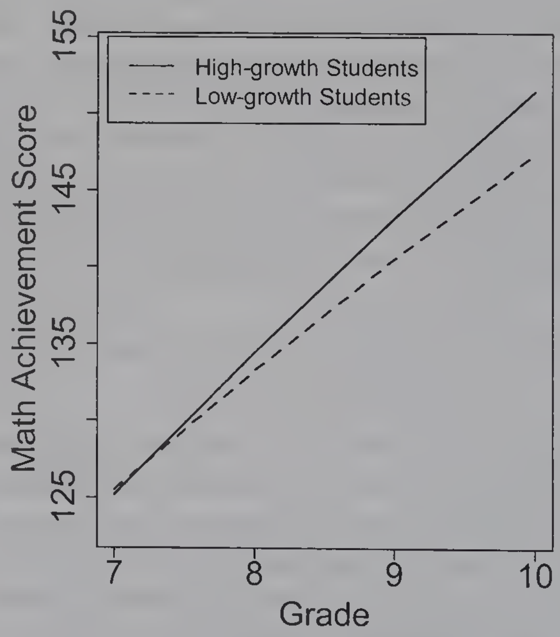


Figure 3. Estimated math achievement growth curve from 7th to 10th grades for high-growth and low-growth students.
Note. To avoid redundancy, only male Gymnasium students are plotted. High-growth students = 1.5 *SD* above the mean for intrinsic motivation and deep learning strategies. Low-growth students = 1.5 *SD* below the mean for intrinsic motivation and deep learning strategies.

achievement in mathematics from Grades 5 to 10. The results showed that motivation and strategies predicted current math achievement over and above intelligence. Furthermore, our findings revealed that motivation and strategies also explained the growth of achievement. Growth was positively predicted by perceived control, intrinsic motivation, and deep learning strategies, and it was negatively predicted by surface learning strategies. In contrast, intelligence turned out not to be a predictor of growth (after controlling for demographic variables), despite being a strong predictor of current math achievement. Overall, these findings largely support our hypotheses.

Relations of Motivation and Strategies With Initial Level and Growth of Achievement

Among the motivational factors, both perceived control and intrinsic motivation predicted long-term growth in math achievement. Specifically, perceived control (assessed at Grade 5) and intrinsic motivation (assessed at Grade 7) positively predicted the total amount of growth in math achievement. This lends support to the notion that these constructs do not merely reflect subjective perceptions of current ability (perceived control) or transient, ephemeral positive emotion (intrinsic motivation), but motivational tendencies that shape future learning and resulting achievement (Deci & Ryan, 1985; Marsh & Craven, 2006). In contrast, extrinsic motivation predicted only initial levels but not growth in achievement, illustrating an interesting contrast with the relations observed for intrinsic motivation. The short-lived nature of extrinsic motivation and long-term benefit of intrinsic motivation are in accordance with previous experimental findings (Murayama & Elliot, 2011; Vansteenkiste et al., 2005), and our results provide the first evidence that this pattern is evident even over long periods of time.

With regard to learning strategies, both deep strategies and surface strategies were related to long-term growth in math achievement, but their relations had opposite signs (deep learning strategies were positively linked, whereas surface learning strategies were negatively linked to growth). Furthermore, surface learning strategies were also negatively linked to initial level of achievement. Although we had not expected that surface learning strategies would lead to lower initial levels and smaller growth rates of math achievement, it is plausible that relying on surface, rehearsal-based strategies can interfere with the use of more efficient strategies and thus be deleterious for achievement outcomes. It should be noted, however, that the present findings do not indicate that surface learning strategies are maladaptive in any situations, as the effectiveness of learning strategies may largely depend on learning materials and phases of learning (Alexander & Murphy, 1998). Surface learning strategies may be useful for some tasks (e.g., skill tasks) or at initial stages of learning.

A seemingly counterintuitive finding is that intrinsic motivation (assessed at Grade 7) and deep learning strategies (assessed at Grades 5 and 7) were unrelated to concurrent achievement. Although unexpected, this is actually consistent with the extant literature suggesting that the relation between these constructs and performance is unclear (Meece et al., 1988; Murphy & Alexander, 2002; Utman, 1997). Students with high intrinsic motivation are less concerned about how well they perform on upcoming achievement tests. Accordingly, although intrinsic motivation should provide long-term benefits, such a noninstrumental approach to learning may not add much to current performance. As for deep, elaborative learning strategies, previous studies indicated that elaborative learning may not be an efficient means of dealing with an upcoming achievement test because semantic elaboration is a relatively slow learning process and therefore costly if time is limited (Garner, 1990). Given the somewhat counterintuitive nature of this phenomenon, further research is needed to replicate the present findings and to examine possible conditions under which these constructs are positively related to (concurrent) achievement scores.

Developmental Differences in the Functions of Motivation and Strategies

It is noteworthy that the motivational and strategy variables predicted growth at different developmental stages. For motivational factors, the relation between perceived control and growth in math achievement was observed when perceived control was assessed at Grade 5, whereas differential relations for intrinsic versus extrinsic motivation emerged only when motivation was assessed at Grade 7.

One possible explanation is that our findings reflect developmental differences in students' understanding of different dimensions of motivation. Perceived control (i.e., cognitive representations of action–outcome contingencies) seems to be a straightforward notion that may require little cognitive capacity to understand. Thus, it is likely that students develop individual appraisals of personal control over achievement and that these appraisals influence their academic achievement from early on (e.g., in the first grades of elementary school). In contrast, understanding the intrinsic versus extrinsic reasons underlying one's behavior is quite complicated and may require higher order thinking. As a result, comprehension of the intrinsic–extrinsic distinction to the extent that intrinsic motivation can clearly relate to long-term achievement growth may occur only later (e.g., in early adolescence). In fact, previous findings indicate that children's perceptions of control (expectancy beliefs) differentiate into subject-specific, distinct constructs at a very early stage in elementary school. In contrast, subcomponents of task value (such as intrinsic value vs. utility value) are less differentiated among younger children, but become more distinct during early adolescence (Wigfield & Eccles, 1992).

We also observed that perceived control at Grade 5 predicted growth in math achievement, whereas perceived control at Grade 7 failed to do so. One possible explanation is that, as mathematics content becomes more difficult, students' judgments of personal control over achievement are anchored more to concrete experiences with current math problems rather than to judgments of personal math ability and possible future trajectories of developing math competence (see Vallacher & Wegner, 1987), resulting in less predictive power for long-term growth. In accord with this interpretation, contrary to the predictive relations with growth, the link between perceived control and concurrent achievement was stronger at Grade 7 than at Grade 5.

With regard to strategy variables, growth in math achievement was positively predicted by deep learning strategies at Grade 7, but not yet at Grade 5. Previous developmental studies on strategy use have shown that students exhibit an increasing readiness across adolescence to deploy strategies that are elaborative and generative (Christopoulos, Rohwer, & Thomas, 1987; Pressley, Borkowski, & Johnson, 1987). In addition, these studies also revealed a "utilization deficiency," a phenomenon whereby children, up until a particular developmental stage, are able to use a learning strategy, but are unable to reap its performance benefits (Bjorklund, Coyle, & Gaultney, 1992; Miller & Seier, 1994). Our findings shed light on these findings, suggesting that an effective use of deep learning strategies may occur only later in students' academic career.

The Role of Intelligence and the Matthew Effect

One of the features of the current investigation is that we controlled for intelligence when examining the predictive relations of motivation and cognitive strategies. This is by itself of considerable importance, as discussed at the outset. In addition, the inclusion of intelligence as a predictor produced interesting findings: Long-term growth in math achievement was predicted by motivational and strategy factors, but not by students' intelligence (after controlling for demographic variables). This stands in marked contrast to the commonly observed finding that intelligence explains a much larger proportion of the variance in current achievement scores, as compared to motivational and strategy variables (e.g., Spinath, Spinath, Harlaar, & Plomin, 2006). We should be aware that this study focused on the development of achievement in one academic domain only. Nonetheless, our findings clearly underscore the importance of paying attention to adolescents' motivation and learning strategies when wanting to understand the development of their academic achievement. Thus, an intriguing message from this study is that the critical determinant of growth in achievement is not how smart you are, but how motivated you are and how you study.

Another interesting finding is the Matthew effect, a phenomenon that over time more able individuals become even better and less able individuals become even worse, thus widening the gap between haves and have-nots. The Matthew effect has been investigated in a number of different domains, such as reading ability (Stanovich, 1986) and job payment (Tang, 1996). The current study observed the Matthew effect in the domain of math achievement during adolescence, and the findings suggest that the effect may have been linked to the German school tracking system. It should be kept in mind, however, that the current analysis provides little information about the background characteristics of students attending schools from different tracks (e.g., in terms of learning environments at home) or the educational quality of these schools that may have produced this Matthew effect. These varied explanations would have different educational implications (see also Baumert et al., 2012). Further research is needed to investigate such mechanisms that widen the gap between students differing in initial achievement.

Limitations and Directions for Future Research

Our findings should be interpreted in the context of several limitations that suggest directions for future research. First, although the IRT scaling has the great advantage of establishing a common metric across different grade levels—an issue that was not addressed in many previous studies—it rests on the critical assumption that all math achievement test items measure a single unidimensional construct. By testing the dimensionality and longitudinal invariance of the PALMA Math Achievement Test, we were able to show that there was no major violation of unidimensionality assumptions in this study (see Supporting Information Appendix S1 available online). However, unidimensionality assumptions cannot be completely met in longitudinal studies of academic achievement, given the need to include items measuring various contents that match the diversity of the curriculum both within and across grade levels. Future research would do well to investigate the impact of such local heterogeneity of items on growth trajectories.

Second, although the present research produced evidence about the predictive power of motivational and strategy variables on growth (after controlling for demographic variables and intelligence), these relations are still correlational in nature. Thus, one should exercise caution when interpreting these links in terms of causality. Although this was not the primary concern of our research, alternative approaches such as a cross-lagged analysis (e.g., Marsh & Craven, 2006) or dual change score analysis (see McArdle & Hamagami, 2001) may be better suited to examine causal ordering and possible reciprocal effects. Previous research has shown that students' academic self-concepts and their achievement can be linked by reciprocal causation (Marsh, 1990). It seems likely that students' intrinsic and extrinsic motivation and surface versus deep strategy use also are reciprocally linked with their academic achievement. Future research should examine the nature of these reciprocal associations.

Conclusion

Numerous studies have been conducted to better understand students' motivation and learning strategies as promoting the academic development of competence and knowledge. Researchers defined motivation as the process whereby goal-directed activity is instigated and sustained (Pintrich & Schunk, 2002). Learning strategies are described as planned sets of coordinated study tactics that are directed by a learning goal and aim to acquire a new skill or gain understanding (Alexander & Murphy, 1998). According to these views, motivation and learning strategies should, by their nature, facilitate long-term learning processes. The present research documents that these variables are indeed important for students' academic growth over the school years.

References

Alexander, P. A., & Murphy, P. K. (1998). Profiling the differences in students' knowledge, interest, and strategic processing. *Journal of Educational Psychology, 90*, 435–447.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Educational Psychology Review, 84*, 191–215.

Bast, J., & Reitsma, P. (1997). Matthew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research, 32*, 135–167.

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development, 83*, 1347–1367.

Bjorklund, D. F., Coyle, T. R., & Gaultney, J. F. (1992). Developmental differences in the acquisition and maintenance of an organizational strategy: Evidence for the utilization deficiency hypothesis. *Journal of Experimental Child Psychology, 54*, 434–448.

Brown, S. C., & Craik, F. I. M. (2000). Encoding and retrieval of information. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 93–107). New York: Oxford University Press.

Butler, R. (1998). Determinants of help seeking: Relations between perceived reasons for classroom help-avoidance and help-seeking behaviors in an experimental context. *Journal of Educational Psychology, 90*, 630–643.

Cervone, D., & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology, 50*, 492–501.

Christopoulos, J. P., Rohwer, W. D., Jr., & Thomas, J. (1987). Grade level differences in students' study activities as a function of course characteristics. *Contemporary Educational Psychology, 12*, 303–323.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*, 5–12.

Enders, C. K. (2006). A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine*, *68*, 427–736.

Entwistle, N. J., & Ramsden, P. (1983). *Understanding student learning*. New York: Nichols.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, *30*, 415–441.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*, 519–533. doi:10.1037/0012-1649.38.4.519

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*, 507–537. doi:10.1111/j.1532-7795.2010.00645.x

Frenzel, A. C., Pekrun, R., Dicke, A. L., & Goetz, T. (in press). Decline of academic interest in adolescence—artifact or real? A study on the nature of change. *Developmental Psychology*.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research*, *60*, 517–529.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, *82*, 525–538.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*, 1357–1371. doi:10.1111/j.1467-8624.2011.01630.x

Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, *52*, 890–898.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology*, *39*, 777–790. doi:10.1037/0012-1649.39.4.777

Hart, D., Hofmann, V., Edelstein, W., & Keller, M. (1997). The relation of childhood personality types to adolescent behavior and development: A longitudinal study of Icelandic children. *Developmental Psychology*, *33*, 195–205.

Heller, K., & Perleth, C. (2000). *Kognitiver Fäigkeitstest für 4. bis 12. Klassen, Revision (KFT 4–12 + R)* [Cognitive abilities test for grades 4 to 12, revision (KFT 4–12 + R)]. Göttingen, Germany: Hogrefe.

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural: Equation Modeling*, *15*, 541–563. doi:10.1080/10705510802338983

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527.

Johnson, W., McGue, M., & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology*, *42*, 514–532. doi:10.1037/0012-1649.42.3.514

Lemos, M. S. (1996). Students' and teachers' goals in the classroom. *Learning and Instruction*, *6*, 151–172. doi:10.1016/0959-4752(95)00031-3

Lerner, R. M. (1993). Early adolescence: Toward an agenda for the integration of research, policy, and intervention. In R. M. Lerner (Ed.), *Early adolescence: Perspectives on research, policy, and intervention* (pp. 1–13).Hillsdale, NJ: Erlbaum.

Lohman, D. F., Ackerman, P. L., Kyllonen, P. C., & Roberts, R. D. (1999). *Minding our p's and q's: On finding relationships between learning and intelligence Learning and individual differences: Process, trait, and content determinants* (pp. 55–76). Washington, DC: American Psychological Association.

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*, 646–656.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*, 133–163.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*, 107–125.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416.

McArdle, J. J., Anderson, E., Birren, J. E., & Schaie, K. W. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 21–44). San Diego, CA: Academic Press.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analysis with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*, 471–490. doi:10.1016/j.ecresq.2006.09.003

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Student's goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, *80*, 514–523.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*, 60–70.

Miller, P. H., & Seier, W. L. (1994). Strategy utilization deficiencies in children: When, where, and why. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 25, pp. 107–156). San Diego, CA: Academic Press.

Murayama, K., & Elliot, A. J. (2011). Achievement motivation and memory: Achievement goals differentially influence immediate and delayed remember-know recognition memory. *Personality and Social Psychology Bulletin, 37*, 1339–1348.

Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *Journal of Experimental Education, 70*, 197–214. doi:10.1080/00220970209599506

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide*. Los Angeles: Author.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.

Otis, N., Grouzet, F. M. E., & Pelletier, L. G. (2005). Latent motivational change in an academic setting: A 3-year longitudinal study. *Journal of Educational Psychology, 97*, 170–183.

Pekrun, R. (1993). Facets of students' academic motivation: A longitudinal expectancy-value approach. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 8, pp. 139–189). Greenwich, CT: JAI Press.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341. doi:10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of quantitative and qualitative research. *Educational Psychologist, 37*, 91–106.

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 17–37). Münster, Germany: Waxmann.

Perry, R. P., Hladkyj, S., Pekrun, R., & Pelletier, S. T. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology, 93*, 776–789.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (2nd ed). Columbus, OH: Merrill-Prentice Hall.

Pressley, M., Borkowski, J. G., & Johnson, C. J. (1987). The development of good strategy use: Imagery and related mnemonic strategies. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processses: Theories, individual differences, and appplications* (pp. 274–301). New York: Springer-Verlag.

Ramsden, P. (1988). Context and strategy: Situational influences on learning. In R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 159–184). New York: Plenum.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill, factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*, 261–288. doi:10.1037/0033-2909.130.2.261

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General & Applied, 80*, 1–28.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749–761.

Shim, S. S., Ryan, A. M., & Anderson, C. J. (2008). Achievement goals and achievement during early adolescence: Examining time-varying predictor and outcome variables in growth-curve analysis. *Journal of Educational Psychology, 100*, 655–671.

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology, 82*, 22–32.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*, 363–374. doi:10.1016/j.intell.2005.11.004

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–406.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–90.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*, 80–90.

Tang, T. L. P. (1996). Pay differentials as a function of rater's sex, money ethic, and job incumbent's sex: A test of the matthew effect. *Journal of Economic Psychology, 17*, 127–144.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107*, 393–419.

Urdan, T., & Midgley, C. (2001). Academic self-handicapping: What we know, what more there is to learn. *Educational Psychology Review, 13*, 115–138.

Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review, 1*, 170–182.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39,* 111–133.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94,* 3–15.

Vansteenkiste, M., Simons, J., Lens, W., Soenens, B., & Matos, L. (2005). Examining the motivational impact of intrinsic versus extrinsic goal framing and autonomy-supportive versus internally controlling communication style on early adolescents' academic achievement. *Child Development, 76,* 483–501.

vom Hofe, R., Kleine, M., Blum, W., & Pekrun, R. (2005). On the role of "Grundvorstellungen" for the development of mathematical literacy–First results of the longitudinal study PALMA. *Mediterranean Journal for Research in Mathematics Education, 4,* 67–84.

vom Hofe, R., Pekrun, R., Kleine, M., & Götz, T. (2002). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA): Konstruktion des Regensburger Mathematikleistungstests für 5.-10. Klassen [Project for the Analysis of Learning and Achievement in Mathematics (PALMA): Development of the regensburg mathematics achievement test for Grades 5 to 10]. *Zeitschrift für Pädagogik, 45*(Beiheft), 83–100.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12,* 265–310.

Woodrow, H. (1946). The ability to learn. *Psychological Review, 53,* 147–158.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACERConQuest Version 2: Generalised item response modelling software.* Camberwell, England: Australian Council for Educational Research.

Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods, 14,* 183–201. doi:10.1037/a0015858

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Additional Analyses.

FELLOWSHIPS AVAILABLE

The **Radcliffe Institute for Advanced Study at Harvard University** awards 50 funded residential fellowships each year designed to support scholars, scientists, artists, and writers of exceptional promise and demonstrated accomplishment.

For more information, please contact:

Radcliffe Application Office
8 Garden Street
Cambridge, MA 02138
617-496-1324
fellowships@radcliffe.harvard.edu
www.radcliffe.harvard.edu

RADCLIFFE INSTITUTE
FOR ADVANCED STUDY
HARVARD UNIVERSITY

Society for Research in Child Development
IS Developmental Science

SRCD Exchange
exchange.srcd.org

SRCD's Social Media Site
Connect with your colleagues in the field!

Top 5 reasons why YOU need to be a SRCD member...

1. Play an Active Role & Broaden Your Horizons in the Society & the Field
2. Link, Share, & Connect with Policy Matters
3. Be "In the Know"
4. Opportunities to Apply for Grants & Awards
5. Many Member Extras

Join SRCD Today!
www.srcd.org

Questions regarding SRCD Membership contact: membership@srcd.org

Save Time and Let the Research Come to You

Sign up for new content alerts for all of your favorite journals:

- ✓ Be the first to read Early View articles
- ✓ Get notified of Accepted Articles when they appear online
- ✓ Receive table of contents details each time a new issue is published
- ✓ Never miss another issue!

Follow these 3 easy steps to register for alerts online:

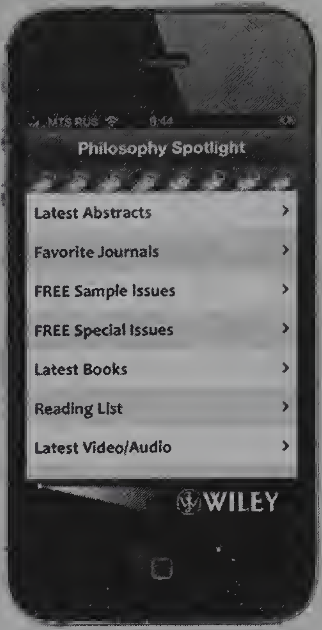
1. Log into **Wiley Online Library**. If you are not already a registered user, you can create your profile for free.
2. Select "**Get New Content Alerts**" from Journal Tools on the top left menu on any journal page or visit the Publications page to view all titles.
3. Submit your preferences and you are done. You will now receive an email when a new issue of the journal publishes.

wileyonlinelibrary.com

L_46050

DOWNLOAD FREE
Wiley Spotlight Apps today!

Whether you want to keep track of broad trends across the discipline or focus in on a subfield, Wiley's subject Spotlight apps are an essential tool for your iPhone or iPad.


MTS RUS
8:44
Philosophy Spotlight
Latest Abstracts
Favorite Journals
FREE Sample Issues
FREE Special Issues
Latest Books
Reading List
Latest Video/Audio
WILEY

Spotlight Apps feature:

- Latest information on key conferences and conference tweets
- Free Frommer's travel guide to major conference locations
- Abstracts for articles and books as they publish with the ability to "follow" your favorites
- Free Publishing Workshops
- Bookmark feature, email a Reading List, and share content via Facebook and Twitter
- Add your own RSS feeds to keep track of your favorite sources

Visit iTunes today to download Spotlight apps in the following subject areas:

Anthropology
Business
Economics
Education
Geography
History
Linguistics
Operations
Philosophy
Politics & Government
Psychology
Religion & Theology
Sociology

WILEY-BLACKWELL

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/publications/child-development.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and ... are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted online at http://mc.manuscriptcentral.com/srcd. Full instructions and support are available on the site and a user ID and password can be obtained on the first visit. Support can be contacted by phone (888-503-1050), or via the red Get Help Now link in the upper right-hand corner of the login screen. If you cannot submit online, please contact the Editorial Office by telephone (734-926-0615) or by e-mail (cdev@srcd.org). Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Copyright Transfer Agreement (http://www.srcd.org/sites/default/files/documents/pubs/cd_cta.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

Full submission guidelines are available at http://www.srcd.org/sites/default/files/documents/submission_guidelines.pdf. In brief, the manuscript file should be formatted with doublespaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (6th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley. com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

RELATIONSHIP PROCESSES AND RESILIENCE IN CHILDREN WITH INCARCERATED PARENTS

- *Julie Poehlmann and J. Mark Eddy* (Serial No. 308, 2013).

THE EMERGENCE OF A TEMPORALLY EXTENDED SELF AND FACTORS THAT CONTRIBUTE TO ITS DEVELOPMENT: FROM THEORETICAL AND EMPIRICAL PERSPECTIVES

- *Mary Lazaridis* (Serial No. 307, 2013).

WHAT MAKES A DIFFERENCE: EARLY HEAD START EVALUATION FINDINGS IN A DEVELOPMENTAL CONTEXT

- *John M. Love, Rachel Chazan-Cohen, Helen Raikes, and Jeanne Brooks-Gunn* (Serial No. 306, 2013).

THE DEVELOPMENT OF MIRROR SELF-RECOGNITION IN DIFFERENT SOCIOCULTURAL CONTEXTS

- *Joscha Kärtner, Heidi Keller, Nandita Chaudhary, and Relindis D. Yovsi* (Serial No. 305, 2012).

"EMOTIONS ARE A WINDOW INTO ONE'S HEART": A QUALITATIVE ANALYSIS OF PARENTAL BELIEFS ABOUT CHILDREN'S EMOTIONS ACROSS THREE ETHNIC GROUPS

- *Alison E. Parker, Amy G. Halberstadt, Julie C. Dunsmore, Greg Townley, Alfred Bryant, Jr., Julie A. Thompson, and Karen S. Beale* (Serial No. 304, 2012).

PHYSIOLOGICAL MEASURES OF EMOTION FROM A DEVELOPMENTAL PERSPECTIVE: STATE OF THE SCIENCE

- *Tracy A. Dennis, Kristin A. Buss, and Paul D. Hastings* (Serial No. 303, 2012).

HOW SOCIALIZATION HAPPENS ON THE GROUND: NARRATIVE PRACTICES AS ALTERNATE SOCIALIZING PATHWAYS IN TAIWANESE AND EUROPEAN-AMERICAN FAMILIES

- *Peggy J. Miller, Heidi Fung, Shumin Lin, Eva Chian-Hui Chen, and Benjamin R. Boldt* (Serial No. 302, 2012).

CHILDREN WITHOUT PERMANENT PARENTS: RESEARCH, PRACTICE, AND POLICY

- *Robert B. McCall, Marinus H. van IJzendoorn, Femmie Juffer, Christina J. Groark, and Victor K. Groza* (Serial No. 301, 2011).

I REMEMBER ME: MNEMONIC SELF-REFERENCE EFFECTS IN PRESCHOOL CHILDREN

- *Josephine Ross, James R. Anderson, and Robin N. Campbell* (Serial No. 300, 2011).

EARLY SOCIAL COGNITION IN THREE CULTURAL CONTEXTS

- *Tara Callaghan, Henrike Moll, Hannes Rakoczy, Felix Warneken, Ulf Liszkowski, Tanya Behne, and Michael Tomasello* (Serial No. 299, 2011).

THE DEVELOPMENT OF AMBIGUOUS FIGURE PERCEPTION

- *Marina C. Wimmer and Martin J. Doherty* (Serial No. 298, 2011).

THE BETTER BEGINNINGS, BETTER FUTURES PROJECT: FINDINGS FROM GRADE 3 TO GRADE 9

- *Ray DeV. Peters et al.* (Serial No. 297, 2010).

FIRST-YEAR MATERNAL EMPLOYMENT AND CHILD DEVELOPMENT IN THE FIRST 7 YEARS

- *Jeanne Brooks-Gunn, Wen-Jui Han, and Jane Waldfogel* (Serial No. 296, 2010).

DEPRIVATION-SPECIFIC PSYCHOLOGICAL PATTERNS: EFFECTS OF INSTITUTIONAL DEPRIVATION

- *Michael Rutter, Edmund J. Sonuga-Barke, Celia Beckett, Jennifer Castle, Jana Kreppner, Robert Kumsta, Wolff Schlotz, Suzanne Stevens, and Christopher A. Bell* (Serial No. 295, 2010).

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET

- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY

- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY

- *Mona El-Sheikh* et al. (Serial No. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN

- *The St. Petersburg–USA Orphanage Research Team* (Serial No. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING

- *Nancy Eisenberg et al.* (Serial No. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE

- *David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS

- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE

- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

WILEY